DOCUMENT RESUME

ED 059 039 SE 013 148

AUTHOR	Jerman, Max
TITLE	Instruction in Problem Solving and an Analysis of Structural Variables That Contribute to Problem-Solving Difficulty.
INSTITUTION	Stanford Univ., Calif. Inst. for Mathematical Studies in Social Science.
REPORT NO	TR-180
PUB DATE	12 Nov 71
NOTE	129p.; Psychology and Education Series
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	*Arithmetic; Computer Assisted Instruction; *Educational Research; *Elementary School Mathematics; Grade 5; *Mathematics Education; *Problem Solving; Programed Instruction

ABSTRACT

This report is divided into two parts. The first part contains the major sections of the author's doctoral dissertation comparing the effects of two instructional problem-solving programs. The fifth grade students in six classes (three schools) were randomly assigned to the two programs: The Productive Thinking Program, a commercially-available sequence which develops general problem-solving skills and contains no mathematics; and the Modified Wanted-Given Program, an experimental sequence which emphasizes the structure of arithmetical problems. Both sequences were presented in programmed form and took 16 consecutive school days. Fifth grade students in two classes in a fourth school acted as a control group. Every student received a pretest, posttest and a follow-up test seven weeks later. Each test battery measured several other skills besides problem solving. On an analysis of covariance, no significant differences were found between the two methods of instruction and the control, nor was any significant sex difference found. The second part of this report reviews the variables used in previous studies of problem solving using teletype terminals, and then applies the same regression techniques to verbal problems selected from the dissertation study described in the first part. (MM)

ED 059039

SE
N-A1

INSTRUCTION IN PROBLEM SOLVING AND AN ANALYSIS OF STRUCTURAL VARIABLES THAT CONTRIBUTE TO PROBLEM-SOLVING DIFFICULTY

BY

MAX JERMAN

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TECHNICAL REPORT NO. 180

NOVEMBER 12, 1971

PSYCHOLOGY & EDUCATION SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

SE 013 148

TECHNICAL REPORTS

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

(Place of publication shown in parentheses; if published title is different from title of Technical Report, this is also shown in parentheses.)

(For reports no. 1 - 44, see Technical Report no. 125.)

50 R. C. Atkinson and R. C. Calfee. Mathematical learning theory. January 2, 1963. (In B. B. Wolman (Ed.), Scientific Psychology. New York: Basic Books, Inc., 1965. Pp. 254-275)

51 P. Suppes, E. Crothers, and R. Weir. Application of mathematical learning theory and linguistic analysis to vowel phoneme matching in Russian words. December 28, 1962.

52 R. C. Atkinson, R. Calfee, G. Sommer, W. Jeffrey and R. Shoemaker. A test of three models for stimulus compounding with children. January 29, 1963. (J. exp. Psychol., 1964, 67, 52-58)

53 E. Crothers. General Markov models for learning with inter-trial forgetting. April 8, 1963.

54 J. L. Myers and R. C. Atkinson. Choice behavior and reward structure. May 24, 1963. (Journal math. Psychol., 1964, 1, 170-203)

55 R. E. Robinson. A set-theoretical approach to empirical meaningfulness of measurement statements. June 10, 1963.

56 E. Crothers, R. Weir and P. Palmer. The role of transcription in the learning of the orthographic representations of Russian sounds. June 17,. 1963.

57 P. Suppes. Problems of optimization in learning a list of simple items. July 22, 1963. (In Maynard W. Shelly, II and Glenn L. Bryan (Eds.), Human Judgments and Optimality. New York: Wiley. 1964. Pp. 116-126)

58 R. C. Atkinson and E. J. Crothers. Theoretical note: all-or-none learning and intertrial forgetting. July 24, 1963.

59 R. C. Calfee. Long-term behavior of rats under probabilistic reinforcement schedules. October 1, 1963.

60 R. C. Atkinson and E. J. Crothers. Tests of acquisition and retention, axioms for paired-associate learning. October 25, 1963. (A comparison of paired-associate learning models having different acquisition and retention axioms, J. math. Psychol., 1964, 1, 285-315)

61 W. J. McGill and J. Gibbon. The general-gamma distribution and reaction times. November 20, 1963. (J. math. Psychol., 1965, 2, 1-18)

62 M. F. Norman. Incremental learning on random trials. December 9, 1963. (J. math. Psychol., 1964, 1, 336-351)

63 P. Suppes. The development of mathematical concepts in children. February 25, 1964. (On the behavioral foundations of mathematical concepts. Monographs of the Society for Research in Child Development, 1965, 30, 60-96)

64 P. Suppes. Mathematical concept formation in children. April 10, 1964. (Amer. Psychologist, 1966, 21, 139-150)

65 R. C. Calfee, R. C. Atkinson, and T. Shelton, Jr. Mathematical models for verbal learning. August 21, 1964. (In N. Wiener and J. P. Schoda (Eds.), Cybernetics of the Nervous System: Progress in Brain Research. Amsterdam, The Netherlands: Elsevier Publishing Co., 1965. Pp. 333-349)

66 L. Keller, M. Cole, C. J. Burke, and W. K. Estes. Paired associate learning with differential rewards. August 20, 1964. (Reward and information values of trial outcomes in paired associate learning. (Psychol. Monogr., 1965, 79, 1-21)

67 M. F. Norman. A probabilistic model for free-responding. December 14, 1964.

68 W. K. Estes and H. A. Taylor. Visual detection in relation to display size and redundancy of critical elements. January 25, 1965, Revised 7-1-65. (Perception and Psychophysics, 1966, 1, 9-16)

69 P. Suppes and J. Donio. Foundations of stimulus-sampling theory for continuous-time processes. February 9, 1965. (J. math. Psychol., 1967, 4, 202-225)

70 R. C. Atkinson and R. A. Kinchla. A learning model for forced-choice detection experiments. February 10, 1965. (Br. J. math stat. Psychol., 1965, 18, 184-206)

71 E. J. Crothers. Presentation orders for items from different categories. March 10, 1965.

72 P. Suppes, G. Groen, and M. Schlag-Rey. Some models for response latency in paired-associates learning. May 5, 1965. (J. math. Psychol., 1966, 3, 99-128)

73 M. V. Levine. The generalization function in the probability learning experiment. June 3, 1965.

74 D. Hansen and T. S. Rodgers. An exploration of psycholinguistic units in initial reading. July 6, 1965.

75 B. C. Arnold. A correlated urn-scheme for a continuum of responses. July 20, 1965.

76 C. Izawa and W. K. Estes. Reinforcement-test sequences in paired-associate learning. August 1, 1965. (Psychol. Reports, 1966, 18, 879-919)

77 S. L. Blehart. Pattern discrimination learning with Rhesus monkeys. September 1, 1965. (Psychol. Reports, 1966, 19, 311-324)

78 J. L. Phillips and R. C. Atkinson. The effects of display size on short-term memory. August 31, 1965.

79 R. C. Atkinson and R. M. Shiffrin. Mathematical models for memory and learning. September 20, 1965.

80 P. Suppes. The psychological foundations of mathematics. October 25, 1965. (Colloques Internationaux du Centre National de la Recherche Scientifique. Editions du Centre National de la Recherche Scientifique. Paris: 1967. Pp. 213-242)

81 P. Suppes. Computer-assisted instruction in the schools: potentialities, problems, prospects. October 29, 1965.

82 R. A. Kinchla, J. Townsend, J. Yellott, Jr., and R. C. Atkinson. Influence of correlated visual cues on auditory signal detection. November 2, 1965. (Perception and Psychophysics, 1966, 1, 67-73)

83 P. Suppes, M. Jerman, and G. Groen. Arithmetic drills and review on a computer-based teletype. November 5, 1965. (Arithmetic Teacher, April 1966, 303-309.

84 P. Suppes and L. Hyman. Concept learning with non-verbal geometrical stimuli. November 15, 1968.

85 P. Holland. A variation on the minimum chi-square test. (J. math. Psychol., 1967, 3, 377-413).

86 P. Suppes. Accelerated program in elementary-school mathematics -- the second year. November 22, 1965. (Psychology in the Schools, 1966, 3, 294-307)

87 P. Lorenzen and F. Binford. Logic as a dialogical game. November 29, 1965.

88 L. Keller, W. J. Thomson, J. R. Tweedy, and R. C. Atkinson. The effects of reinforcement interval on the acquisition of paired-associate responses. December 10, 1965. (J. exp. Psychol., 1967, 73, 268-277)

89 J. I. Yellott, Jr. Some effects on noncontingent success in human probability learning. December 15, 1965.

90 P. Suppes and G. Groen. Some counting models for first-grade performance data on simple addition facts. January 14, 1966. (In J. M. Scandura (Ed.), Research in Mathematics Education. Washington, D. C.: NCTM, 1967. Pp. 35-43.

91 P. Suppes. Information processing and choice behavior. January 31, 1966.

92 G. Groen and R. C. Atkinson. Models for optimizing the learning process. February 11, 1966. (Psychol. Bulletin, 1966, 66, 309-320)

93 R. C. Atkinson and D. Hansen. Computer-assisted instruction in initial reading: Stanford project. March 17, 1966. (Reading Research Quarterly, 1966, 2, 5-25)

94 P. Suppes. Probabilistic inference and the concept of total evidence. March 23, 1966. (In J. Hintikka and P. Suppes (Eds.), Aspects of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1966. Pp. 49-65.

95 P. Suppes. The axiomatic method in high-school mathematics. April 12, 1966. (The Role of Axiomatics and Problem Solving in Mathematics. The Conference Board of the Mathematical Sciences, Washington, D. C. Ginn and Co., 1966. Pp. 69-76.

(Continued on inside back cover)

INSTRUCTION IN PROBLEM SOLVING AND AN ANALYSIS OF STRUCTURAL VARIABLES THAT CONTRIBUTE TO PROBLEM-SOLVING DIFFICULTY

by

Max Jerman

TECHNICAL REPORT NO. 180

November 12, 1971

PSYCHOLOGY AND EDUCATION SERIES

Reproduction in Whole or in Part is Permitted for

any Purpose of the United States Government

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

PREFACE

The problem of how children learn to solve verbal problems in arithmetic has been a subject of research by the Institute and the author for the past several years. All the research on problem solving at the Institute to date has been conducted in the context of computer-assisted instruction (CAI). This report discusses problem solving using traditional paper-and-pencil methods, which was part of the author's doctoral dissertation, and includes further regression analysis of the same data base using variables from previous studies in which students at teletype terminals served as subjects. One purpose for further analysis was to determine if the variables previously found to account for a large proportion of the variance in a CAI setting would also account for a large proportion of the variance in the traditional paper-and-pencil problem-solving setting.

The first four chapters contain the major sections of my dissertation. Chapter V presents a review of the variables used in previous studies of problem solving with students at teletype terminals. Chapter VI presents the results of analysis on verbal problems selected from the dissertation study, using the regression techniques developed in earlier work. A comparison of the goodness of fit of the model is then made for each method.

I would like to acknowledge the interest and assistance of Dr. E. G. Begle for his direction and guidance and for making computer time available through the School of Mathematics Study Group to analyze the data reported in Chapters I-IV. Mr Ray Rees of SMSG gave much of his time processing the

data, while indispensable friends, namely, Mrs. Arlene Dyre, Mrs. Velma Hoffer, and Mrs. Grace Kanz aided in administering the various treatments and tests used in this study.

I am also grateful to Professor Patrick Suppes of IMSSS for providing needed facilities, including the Institute's PDP-10 computer services, materials and for partially supporting this study with funds from National Science Foundation Research Grants NSFG-18709 and NSFGJ-433X. The data for Chapters V-VI were processed by the author.

Finally, I appreciate the assistance, patience, and encouragement of my committee and especially my wife, Roberta.

Table of Contents

Preface i

Table of Contents iii

Chapter

I. Introduction 1
The Problem 3
Review of Research on Problem Solving 4
Summary and Hypotheses 43

II. Experimental Design 45

III. Results 62

IV. Summary and Discussion 79

References 116

List of Appendices 121

CHAPTER I

INTRODUCTION

One of the most important goals of mathematics instruction is to develop in students the ability to solve verbal problems (Kramer, 1966, p. 349). For the early grades, textbook writers and mathematicians recommend that primary emphasis be placed on understanding the problem and that secondary emphasis be placed on computing the answer (Goals, 1963, p. 36). For the later grades, they recommend that emphasis be placed on establishing a set of basic rules such as those that follow, for use in problem solving (Duncan, Capps, Dolcioni, Quest, 1967, p. 54):

1. Identifying the sets involved;
2. Determining whether the sets are to be joined, separated, or compared;
3. Writing an equation that corresponds to the set operation;
4. Solving the equation;
5. Interpreting the solution in terms of the sets involved.

The hope is that by applying the techniques embodied in such rules the students will be able to transfer the problem-solving skills learned in the classroom to real-life situations in later life. In most cases, however, the set of rules, together with the problem set to which they are applied, follows chapters on specific topics in the textbooks. Rather than creating a true problem-solving situation, the exercises often represent little more than a verbal application of the computational skills introduced in a preceding chapter.

Traditionally, instruction in problem solving either centers around teaching students to follow some set of rules, steps, or heuristics, or simply eliminating all rules: "The best way to teach children how to solve problems is to give them lots of problems to solve (Van Engen, 1959, p. 74)." One popular text instructs its seventh-grade readers to solve problems as follows (Eicholz, O'Daffer, Brumfield, Shanks, Fleenor, 1967, p. 131):

> In working story problems, it is helpful to be able to compute rapidly and accurately. It is even more important to be able to decide what to do with the numbers given in the problem. It is impossible to memorize rules that will tell you how to work every problem you may need to solve. You simply must think carefully about the information given and then decide what operations to perform upon the numbers.

The text continues with several sample solutions to problems with the words "The following examples show how you might reason in order to solve difficult problems."

Two new approaches to teaching problem solving have been advanced in recent years. One was a modification of the traditional "wanted-given" approach (Wilson, 1964). The other emphasized problem solving at a general level, that is, it was not oriented to any particular academic discipline (Covington, Crutchfield, and Davies, 1966). Results from each of the programs indicated that students who completed the respective programs made significantly greater gains on posttest measures of creative thinking or problem solving than did their respective control groups.

The question examined in this study was whether students who had received training in general problem-solving skills by using The Productive Thinking Program Series One: General Problem Solving (Covington, Crutchfield, and Davies, 1966) would achieve significantly higher scores on posttest measures of problem solving in mathematics than would students who had received specific training in problem solving in mathematics using a Modified Wanted-Given Program approach.

THE PROBLEM

The purpose of this study was to investigate the differential effects of two instructional programs on performance of verbal mathematical problems by students at the fifth-grade level. In previous studies using one of the programs, The Productive Thinking Program, student performance consistently improved on tests of creative thinking and general problem solving. In the other program a modification of a program developed by Wilson (1964), students were instructed to solve problems in mathematics using a Modified Wanted-Given approach. The central question was whether a general approach to problem solving, such as The Productive Thinking Program, would produce higher performance on criterion tasks, primarily mathematical problem-solving tasks, than would a program which used a mathematical context to teach problem solving in mathematics.

For some students, all problems are not really problems and calling a set of sentences a problem is somewhat arbitrary. As Cronbach (1948) pointed out "a situation presents a problem only when one must give a response (that is, when he seeks satisfaction) and has no habitual response which will give satisfaction. [p. 32]" The term "problem" as used in this study refers to a statement in written form that requires a written response. This definition is in agreement with that given by a standard reference in the field, the Mathematics Dictionary (James and James, 1968), "a question proposed for solution; a matter for examination; a proposition requiring an operation to be performed or a construction to be made, as to bisect an angle or find an eighth root of 2. [p. 286]" The distinction between an exercise and a problem noted by Henderson and Pingry (1953) was observed in the course of the study. Their definition is as follows:

> With the exception of the syntactical form, the chief difference between exercises and verbal 'problems' lies in their intended use. Exercises, such as those dealing with the fundamental operations, exponents, radicals, the binomial theorem, and derivatives, are for the purpose of teaching certain mathematical concepts and generalizations. Verbal problems are for the purpose of teaching the generalizations relative to the process or method of problem solving. These have no necessary relation to a particular kind of mathematics problem; the problem-solving process is essentially the same for all problems [p. 235].

Throughout the study the emphasis in the problem-solving programs was on the process rather than the teaching of certain specific mathematical concepts. The terms "word problem," "stated problem," "story problem," or simply "problem" were all used synonymously.

REVIEW OF RESEARCH ON PROBLEM SOLVING

Research on problem solving in elementary-school mathematics has not been systematic and results often conflict. This brief review of previous studies includes only the major areas in which studies have been conducted to give the reader a feeling for the diversity of studies in problem solving in mathematics and to show how this study is related to past work in the field of problem solving in mathematics. Initially the review includes studies that used subjects at different age levels. Later, the review focuses on students in the upper elementary grades, e.g., grades 4, 5, and 6. The studies are categorized, generally, according to the classification developed by Kilpatrick (1969).

Problem-solving Ability

Individual differences. Dodson (1970, p. 104) prepared a composite list of the strongest characteristics of a good problem solver in mathematics. Basing his evaluation on data involving approximately 1,500 tenth- through twelfth-grade students from the National Longitudinal Study of Mathematical Abilities he concluded that a good problem solver:

1. performs higher on all of the mathematics achievement tests than the poorer problem solvers.
2. performs high in solving mathematics problems that require a great deal of synthesis.
3. solves algebraic equations proficiently.
4. performs well on more advanced mathematics achievement tests administered a year after the criterion test.
5. scores high on verbal and general reasoning tests.
6. determines spatial relationships successfully.
7. resists distraction, identifies critical elements, and remains independent of irrelevant elements.
8. is a divergent thinker.
9. has low debilitating test anxiety while his facilitating test anxiety remains high.
10. has a positive attitude toward mathematics.
11. sees himself as a good mathematics student and does not wish to be a better mathematics student.
12. has teachers who had the most credits beyond the bachelor's degree.

13. has teachers who had the highest degrees.
14. comes from a family with a relatively high income.
15. comes from a community in which the starting teacher's salary is higher than the starting salary of the teachers of his poorer problem-solving peers.
16. lives in a community that has had a recent change in the size of the population.
17. has a socioeconomic index that was about the same as that of a poorer problem solver.

Tate and Stanier (1964) found junior high school students differed in their ability to judge when they had enough facts to solve a problem. Poor problem solvers tended to select answers that had a high affective component on the practical judgment test.

Sex differences in the problem-solving ability of ninth-grade students were studied by Sheehan (1968). After five weeks of special treatment, small-step, linearly programmed materials and problem units accompanied by reference texts, significant differences in the adjusted scores on a problem-solving task in algebra favored the boys. Sheehan concluded that sex did make a difference in learning to solve problems in algebra and other school learning as well.

Few definitive studies on individual differences in problem-solving ability of students in grades 4, 5, and 6 have been made.

Related skills and abilities. In a study using 1,400 sixth graders, Balow (1964) established that both reading ability and computation ability correlated with problem-solving ability as measured by the arithmetic reasoning scale of the Stanford Achievement Test. He found no interaction between computational and reading ability when I.Q. was controlled. Comparing the F-ratios Balow concluded that computation is a much more important factor in problem solving than reading ability. A year earlier, however, Martin (1963) noted that the partial correlation between reading and problem solving with computation held constant was higher at both fourth- and eighth-grade levels than the partial correlation between computation and problem solving with reading held constant.

Wederlin (1966) rotated two factorial studies to a common structure and identified five factors common to both studies on which the loadings were almost identical. He concluded that problem solving in mathematics depends primarily on a general reasoning factor. The deductive reasoning and numerical factors were held to be of somewhat lesser importance. Wederlin determined that little is known concerning the nature of deductive reasoning and numerical factors. The majority of the 371 subjects in both of Werdelin's studies were 13, 14, and 15 years old. Very (1967) found four reasoning factors, arithmetic, deductive, inductive, and general for males in his study involving 355 college students. By performing separate factor analyses by sex, he affirmed that males, in general, tended to have a greater number of mathematical abilities than did the females. Further, the abilities of the males were more specific and more easily identified. A later study (Dye and Very, 1968) revealed ninth-grade females were generally superior to males in tests of numerical facility and perceptual speed. At the eleventh-grade level the difference in favor of females was somewhat less. At the college level males were superior in every test of arithmetic reasoning and mathematics aptitude while females retained superiority in perceptual speed and frequently demonstrated superiority in verbal skills.

Affective variables. A study involving 358 sixth graders by Jonsson (1965) reported some interaction between test anxiety and test difficulty, expecially for girls. Other results of the study suggested that anxiety may act to distort the findings in studies of problem solving, particularly for difficult and complex problems. Difficult tests may have a detrimental effect on the performance of highly anxious students.

In a study conducted by Gangler (1967), college students who were told their work on a problem-solving task in logic would count toward their grade in mathematics achieved lower scores on the task than those who were not told. The negative effect was greatest for high I.Q. students. One of the most important conclusions of this study was that overt participation in the learning situation apparently made the students more flexible in problem solving and resulted in fewer errors. It was also summarized that the type of participation, overt or covert, as well as motivation were important in problem solving and varied according with the students.

Low-ability students in two English schools were less frustrated in a low-competitive school environment and were also less rigid in their approach to solving problems than comparable children in a traditional school (Kellmer, Pringle, and McKenzie, 1965).

Problem-solving Tasks

There is such variety in the types of problems used in various studies that generalizations of the findings of any one study seems unwarranted (Kilpatrick, 1969). Ray (1955) proposed a set of 29 problems with dimensionable independent variables that he hoped would become a standard set of materials to use in studies of human problem solving. Eleven years later, Davis (1966) summarized the status of research and theory in human problem solving by considering one category of tasks as those that required covert trial-and-error behavior and the other category as those that approach overt trial-and-error. To date, standardized sets of problems have not had any wide-scale use. The following categories include most of the types employed by contemporary studies.

Problem content. Studies of problem structure have examined variation in the language used in stating a word problem such as the presence or absence of an existential quantifier, the amount of redundancy, selectivity, and contiguity in problem statements (West and Loree, 1968), and placement of the question in the problem statement (Williams and McCreight, 1965). The results indicated that curriculum developers should insure a variety of problem statements in which they use different words to describe situations. Generally, problems with a high degree of selectivity (little irrelevant data) and redundancy (repeating the data) are easier for both seventh- and ninth-grade students. Decreasing the contiguity (increasing the distance between data) makes problems more difficult for seventh graders, but not for ninth graders. Asking the question first in a word-problem statement may be more helpful for some fifth- and sixth-grade students, even though the differences may not be significant in terms of improved performance. Problems are apparently easier for fifth-grade students to solve when the numerical data are presented in order of use (Burns and Yonally, 1964) rather than in mixed order. Students with low arithmetic reasoning ability find problems in mixed order more difficult. It was suggested that students be given problems with proper order in the initial stages of instruction. The results of a study by Stull (1964) to determine the effect of auditory reading assistance suggested that problem difficulty is more a function of reasoning ability than reading skill for groups of fourth-, fifth-, and sixth-grade students. The readability of problems was found to interact significantly with mental ability in a study by Thompson (1967). The data suggested that mental ability and readability are each significant in their effect on problem solving. The effect of readability was greatest for those of low mental ability.

Problem-solving Processes

Developmental changes. The work of Piaget has generated many studies in recent years. Several of the studies demonstrated that young children can learn and transfer quite complicated problem-solving strategies (Stern, 1967; Stern and Keislar, 1967; Wittrock, 1967).

Studies using problems in elementary sentential logic with first, second, and third graders (O'Brien and Shapiro, 1968; Hill, 1960) have demonstrated that there are differences in the ability of young children to recognize logically necessary conclusions and their ability to test the logically necessary conclusion. While no evidence of growth in ability to test logical necessity was apparent, there was evidence of corresponding growth in ability to recognize logical necessity.

Strategies of inquiry. To date, nearly all studies in this area have been concerned with concept attainment and have used non-mathematical tasks. Neimark and Lewis (1967) suggested that the acquisition of information-gathering strategies is essentially all-or-none for each individual. The evidence which indicates an increase in the use of information gathering by older students may be interpreted as simply meaning that more individuals acquire strategies with age.

Heuristic methods. Polya (1957) generated a great interest in heuristic methods with his set of rules for solving problems. Several studies on problem solving have attempted either to teach heuristic methods or to analyze problem-solving protocols using a system based on Polya's system or various other techniques. Wilson (1967) stated, in a study involving instruction of algebraic concepts to high school students, that instruction should begin with presentation of the content, be followed by task-specific heuristics and then be concluded with general problem-solving heuristics. Kilpatrick (1969) noted that in many cases the important factor is how the problem is seen by the student. He observed that the important question of how a student adapts various heuristics to different kinds of problems is yet to be studied in any depth. Zweng (1968, p. 253) went a step further when she said "We need to know more about how children do, in fact 'see' the physical world."

Creative Thinking

Creativity and problem solving have been the subjects of a number of studies, but few have dealt directly with problem solving in mathematics. Tests developed to measure creativity or to identify creative students sometimes proved disappointing (Prouse, 1967), while others looked promising (Evans, 1965), even though more data on their validity and reliability are needed before they can be used on a broad scale.

The results of a study by Klein and Kellner (1967) involving 130 male, undergraduate students on a two-choice probability-learning task seemed to indicate that highly creative students demonstrate a greater tendency to form hypotheses about patterns of reinforcements than students who are not as creative. Though not working with mathematical tasks, Eisenstadt (1966) found creative college students faster at solving puzzle problems. The approach used by creative students was apparently different from the approach used by noncreative students. Earlier, Mendelsohn and Griswold (1964) reported a positive relationship between creativity and the use of incidental cues in solving anagram problems.

A natural question to ask at this point is whether it is possible to train people to think creatively and to transfer the skills learned to solve problems in mathematics.

Summary

The diverse nature of the majority of studies in problem solving in mathematics is evident in the partial list of categories reviewed above. The subjects who participated in these studies varied in age from 5 years to college age. Taken as a whole, these studies offer only general recommendations of how to approach the problem of teaching fifth-grade students to solve problems in mathematics. A summary of some specific recommendations is given in a later section.

The work by Covington, Crutchfield and Davies (1966), to be described in the following section, represents one new approach that is worthy of investigation. Their program has apparently produced significant gains on tests of creative thinking and problem solving in some cases, but not in others. Should their program prove to make a significant contribution to the problem-solving ability of fifth-grade students in mathematics, considerable support could be mustered for their approach, the theory and assumptions behind it, and perhaps a new approach to the improvement of instruction in problem solving in the upper elementary school could be designed.

A second approach to teaching problem solving in mathematics at the fifth-grade level is one which employs a modified wanted-given approach based on the work of Wilson (1964) and others. This approach provides instruction in problem solving in a mathematical context rather than the more general approach used by the program developed by Covington and others. Each of these programs is described in detail in the following sections.

The Productive Thinking Program: Assumptions. The program was designed to strengthen the elementary-school student's ability to think and it is an outgrowth of research on productive thinking conducted by Professors Richard S. Crutchfield and Martin V. Covington and colleagues in the Department of Psychology and the Institute of Personality Assessment and Research at the University of California, Berkeley. Crutchfield (1966) stated some causes and conditions that seem to contribute to poor problem-solving behavior by most elementary-school students.

> In short, the child in the all-too-typical school situation is being expected to develop efficient cognitive skills, such as those in problem-solving, under conditions where he is offered few opportunities for actual practice of the skill, where the practice he does get is likely to consist of tasks that are too easy, too repetitive, and seem meaningles and trivial to him, where he is often rewarded for low-level performance on these tasks, where he can often just passively listen instead of actively trying out the skill, where he gets incomplete and delayed evaluative information about how well or poorly he is doing and little specific indication of just what he is doing right and what wrong. It is clear that no respectable athletic skill could be developed under such bizarre handicaps (nor would any coach countenance them), and it is doubtful that any complex cognitive skill could be so developed [p. 65].

To overcome the current inadequacies in teaching cognitive skills, Crutchfield recommended five instructional steps to remedy the situation (Crutchfield, 1966, pp. 65-66). They are:

> 1. In the sweeping reconstruction of curriculum materials currently in progress, a good deal of emphasis should be put on designing the materials deliberately in such a way as to demand the exercise of complex cognitive skills of problem-solving and cognitive thinking to a degree compatible with other curricular aims. . . .
>
> 2. Once such curriculum materials are available, they should be studied as they are tried out in the schools, with particular emphasis on a detailed observational analysis and evaluation of children's performances in the various skills. . . .
>
> 3. Better teaching techniques for the fostering of cognitive skills must be developed. In particular, the child should be helped to identify, discriminate, and understand the nature and function of the skills involved in various kinds of cognitive tasks. . . .
>
> 4. In curriculum materials and instructional efforts great stress should be placed on the transfer of cognitive skills. The child should be brought to understand the wide applicability of these complex skills to other subject-matter problems and to other fields. . . .
>
> 5. Finally, attention must be given to the development and refinement in the child of an indispensable, overmastering cognitive skill--the skill of organizing and managing the many specific cognitive skills and resources one possesses for effectively attacking a problem. It is in part the possession of such a master skill that distinguishes the truly productive thinker and creator from the merely talented person. . . .

The Productive Thinking Program, then, is aimed at promoting the generalized problem-solving skills of fifth- and sixth-grade students "skills which are likely to be applicable to a wide range of problems in many different subject-matter fields" (Crutchfield, 1966, p. 68).

The instructional materials themselves have been published as programmed booklets using a comic-book format. Students follow the story-line activities of two children as they encounter and learn to solve problems in a detective-like fashion under the direction of their uncle. Jim and Lila, a brother and sister, are intended to be models with whom students identify as they read the stories and answer the questions in the program. The materials were written in programmed-text format to accomplish several purposes. These purposes are (Crutchfield, 1966, p. 69):

> 1. To demonstrate and emphasize in problem-solving the value, necessity, and techniques of identifying and defining a problem properly, of asking questions and taking time for reflection rather than leaping to conclusions, of looking carefully at details and searching for discrepancies, of generating many ideas and not fearing to come up with what may seem 'silly' ideas, of looking everywhere when stymied for possible clues and sources of ideas, etc.
>
> 2. To give the child an opportunity to generate his own ideas and to become more familiar with his own thought processes and individual cognitive style.
>
> 3. To give the child immediate and frequent feedback to his ideas in the form of good examples of fruitful questions and hypotheses, either as confirmation of his own ideas or as helpful guides to his thinking.
>
> 4. To allow the child to 'participate' in the solution of problems with a pair of curious and imaginative children who serve as identification-models.
>
> The training materials have been written in a free adaptation of the programed-instructional form in an effort to capitalize on several crucial advantages of this method: (1) individual self-administration, permitting freedom from direct group and teacher pressures, permitting individual self-pacing and freedom from interruption of one's own thought, etc; (2) immediate feedback for each child contingent upon his own individual responses; (3) the requiring of active involvement by the child in the materials; (4) the providing of greater scope for exploitation of and accomodation to the divergent and idiosyncratic responses of the child; (5) built-in diagnostic tests or 'indicators' of the sequential progress of the child's thinking, of the specific difficulties being encountered, etc. Such book-form programmed materials can be administered in a flexible manner by the teacher as a supplementary part of other classwork, requiring little or no direct intervention by the teacher and no special training of the teacher in the use of the materials.

In summary, the materials were designed to develop and strengthen a student's ability in using important skills and strategies for thinking and problem solving, to improve a student's awareness of his own thinking processes, and to improve his attitude toward activities that involve use of the mind.

The first main study using the program was conducted in the spring of 1963 (Covington and Crutchfield, 1965) with two pairs, experimental and control, of fifth-grade classes and one pair of sixth-grade classes, 195 students in all. A six-hour pretest of creative thinking tasks was given to all children. Students in the experimental classes were given 13 one-hour lessons in the training program while students in the control classes were given some stories to read and questions to answer the last few days of the treatment period before the posttest. The posttest was an eight-hour battery. Five months later a one-hour follow-up test was given to as many of the fifth-grade students as they were able to locate. It was reported that ". . . the 98 instructed children markedly out-performed the 97 control children" (p. 4) on almost all of the problem-solving tests on the criterion battery. The marked superiority of the instructed children was found in measures of divergent thinking and originality. The instructed students also showed significant positive changes in the degree to which they valued problem solving. The instructed children were superior in achievement on the follow-up test compared with the control students. It was concluded that the training had effects on a considerable variety of problem-solving tasks and that the effects continued for some time during which no instruction was given. Other results indicated that boys and girls made approximately equal gains on posttest measures following training.

After the favorable findings of the first study the training program was revised and expanded to 16 lessons. Two internal criterion tests were also included in the new program "to provide data on how rapidly the superiority of the instructed children over the control children develops (p. 5)." A second version of the program (Passive Exposure) was prepared with a slightly reduced response requirement. A third treatment provided children with only a set of rules to aid their thinking. These rules were presented immediately prior to each of the two internal criterion tests and at the start of the posttest sequence. This second study was conducted with 286 fifth- and sixth-grade students. In general, the results corroborated the findings of the 1963 study. The Passive Exposure group demonstrated a satisfactory degree of proficiency on the posttest and it was concluded that the Passive Exposure condition was equally effective with the regular, programmed version of the treatment. The Rules-Only group placed the lowest of the treatment groups on the posttest criterion measures for the fifth grade, but still higher than the controls. The program was found to be equally effective with both boys and girls. Rather than interpret the findings as evidence that new problem-solving skills had been instilled, Crutchfield concluded that the training might act to "sensitize the child to use the skills he already possesses (p. 5)."

Perhaps the most up-to-date summary of the assumptions about teaching creative thinking on which Covington, Crutchfield, and Davies based the latest version of their program is the following given by Olton

(1969), who assisted with the development of test materials for a large-scale test of the program.

> First, we assumed that virtually all students, regardless of intelligence or initial level of competence, demonstrate a level of thinking that falls far short of what they are potentially capable, and that appropriate instructional materials could bring about a substantial increase in the extent to which a student utilizes his potential for creative thinking.
>
> Second, we assumed that the skills involved in creative thinking are general skills--that is, they cut across curriculum boundaries. They are general cognitive abilities, such as the production of original ideas, the invention of a unifying principle which integrates several disparate events, and the use of various strategies when one is 'stuck' on a complex problem.
>
> Third, we felt that the facilitation of creative thinking could be accomplished without making major changes in the basic cognitive capacities of the student. Instead our instructional efforts would seek to develop, strengthen, and integrate skills and attitudes which the student already possessed in some measure, rather than attempting to develop entirely new and basically different cognitive capacities.
>
> Finally, although we are in favor of teaching traditional subject matter in ways that promote creativity, we felt that direct training of productive thinking skills, in addition to imaginative teaching of curriculum material, would be more effective than either of these techniques alone.

The most exensive test of the revised program to date was conducted at Racine, Wisconsin in 1966. This study was designed to test "the instructional limits of the materials by using them as an entirely self-contained program, with all forms of teacher particiapation purposely held to a minimum." (Olton et al., 1967, p. 31). All students in 44 of the 47 fifth-grade classes in the Racine Unified School District No. 1 were subjects for the experiment. Classes were rank ordered along a facilitative or nonfacilitative scale by a school district supervisor. A "facilitative environment" meant a room atmosphere in which topics were discussed in a "free, creative interchange between teacher and students (p. 3)" Pretest and posttest batteries were developed by the Creative Thinking Project at Berkeley and by Dr. E. Paul Torrance of the University of Minnesota. None of the batteries included mathematical problems.

Of the 21 analyses of pretest variables, only one treatment effect was significant at the .05 level. The treatment group achieved a significantly greater number of solutions to the Jewel problem, but did not achieve a correspondingly greater number of ideas or produce a higher quality of ideas than did the control classes. On the posttest, however, the treatment group scored higher ($P < .05$ or .01) on 12 variables while the control scored higher ($p < .05$) on only one variable. The superiority of the treatment group was apparent at all three levels of I.Q. identified for the purposes of the study. In the single case on which the control group scored significantly higher than the experimental group on the posttest it was found that only 15 percent of all students were able to solve the problem (the X-ray problem). Due to the relatively small number of students who were able to solve the problem, it was suggested that the superiority of the control group was an artifact of the problem itself (p. 21). The treatment seemed to be more effective with students in nonfacilitative environments as evidenced by the reduction of initial differences between facilitative and nonfacilitative classes on the posttest measures compared with pretest measures. The reverse was true for the control group. No significant Treatment by Sex interactions indicated that this version of the program was also equally effective with both boys and girls. Olton et al. (1967) summarized that:

> A diverse set of performance indicators, each reflecting a different aspect of the total problem-solving process, showed consistent benefits as a result of training. These included achievement of solution to problems, number and quality of ideas produced, intellectual persistence, sensitivity to discrepant or puzzling facts, use of a Master Thinking Skill, and an understanding of the process of thinking itself [p. 31].

Using eighth-grade students as subjects, Ripple and Dacey (1967) found evidence to support the transferability of direct training in generalized problem solving to another task, the two-string problem. The treatment used in their study was a modified version of The Productive Thinking Program, upgraded to an eighth-grade level by eliminating some of the repetition in the program and adjusting the vocabulary level. The students who received the special treatment were able to solve the problems given in the posttests significantly faster than their controls. No significant differences were found between experimental and control groups on measures of creative thinking. It was noted in this study that the instructional effects were less potent than those reported by Covington and Crutchfield (1965) for grades 5 and 6. The effects of their treatment were somewhat less for sixth graders than for fifth graders. It may be that the optimum grade level for which instruction in problem solving using material prepared in this format is grade 5. Torrance (1967) referred to a phenomenon known as the fourth-grade slump in creative thinking. His studies showed that after grade 5 there is a general lowering in the level of achievement on tests of creative thinking in several cultures, including our own. He identified several periods corresponding to grades K or 1, 5, 7, and 12 during which students do not perform as highly as expected on tests of creative thinking. It has been suggested (Torrance, 1967; Ripple and Dacy, 1967) that The Productive Thinking Program be revised for grades 6 and beyond in an attempt to overcome this problem.

Olton (1969) listed the areas of attitude and motivation as two of the most important gains produced by The Productive Thinking Program. He cited evidence to show that fifth-grade students not only did improve attitudes toward problem solving, but that the improvement prevailed on a follow-up test more than six months later.

A study by Treffinger and Ripple (1969) failed to provide evidence for the transfer from the instructional materials to problem solving on a specially prepared set of arithmetic problems for students in grades 4 through 7. It was acknowledged, however, that the tests were difficult and their reliability was lower than desirable. When I.Q. was controlled, there were no significant differences between control and experimental pupils' mean scores on the verbal creativity tests. Evidence for transfer from the instructional materials to English was cited by Olton and Crutchfield (1969). Essays written by students who had completed The Productive Thinking Program were judged more thoughtful than those written by a control group. Also, the essays from the experimental group were judged superior in the amount and quality of thinking.

In the above studies, no real test has been made of transfer of the skills taught in The Productive Thinking Program to problem solving in mathematics in which the majority of problems required more than a single step. Olten et al. (1967) noted a lack of treatment effects on many of the performance measures for the brief problems, those which emphasized divergent thinking. Although no examples of the arithmetic problems used in the Treffinger and Ripple study (1969) were given, the titles "A Puzzle Form and a Text Problems Format" indicate that brief problems of at most two steps were probably used since a survey of the mathematics texts used in grades 4 through 7 furnishes abundant evidence of consistent lack of even two-step problems. As stated earlier, most "problems" given in most mathematics texts are simply "exercises" or applications rather than problems that require serious thought on the part of the student. The training provided by The Productive Thinking Program should be most evident in situations where the solution to the given problem requires two or more steps. In short, the transfer effect of The Productive Thinking Program to problem solving in mathematics has not been tested with multiple-step problems. One purpose of this study was to investigate the effectiveness of The Productive Thinking Program on multiple-step verbal problems in arithmetic.

An important factor in the effectiveness of The Productive Thinking Program is the amount of teacher-directed discussion of each lesson. Covington and Crutchfield (1965) and Olton et al. (1967) found that students achieved up to 50 percent higher scores on posttest measures in classes where the teacher discussed each lesson. When teachers participated by discussing the lessons with students, new variables entered that made it difficult to examine the possible effects of the treatment itself. As Olton et al. (1967) pointed out, the severest test of the effectiveness of a set of instructional materials is its use on a daily basis without teacher participation. It was under these carefully controlled conditions that the performance of the experimental classes in the Olton (1967) study surpassed that of the control group on 30 of the 40 internal and posttest measures. Although only 13 of the differences found were significant, 11 favored the experimental group.

The Productive Thinking Program has demonstrated its effectiveness in instruction in general problem solving for the more complex types of problems. Follow-up studies show that students retain this superiority over their controls for periods of 5 months or longer. Since most verbal problems in mathematics texts can be classified as exercises rather than problems, and one of the goals of instruction in mathematics is to teach students how to solve complex problems, I believe that The Productive Thinking Program was worthy of being used as one of the treatments in this study to determine its transfer effect to problem solving in mathematics when problems of a more complex nature than usual, for the grade level, were used as criterion measures.

We now turn to the consideration of the other treatment and the premises upon which it was based.

The Modified Wanted-Given Program. The task of how to combine the most promising suggestions of previous research in problem solving in such a way as to derive an instructional program was not at all clear-cut. The findings of many of the studies which reported significance, in one way or another, were in doubt because of their poor experimental design. Kilpatrick (1969, p. 179), in reviewing some 117 studies in problem solving at all grade levels, noted an increase in the complexity of design in studies during the last few years. He suggested that more clinical studies with individual students should be conducted before beginning larger studies because:

> . . .unfortunately, the increasing complexity of design has been accompanied by an increasing number of methodological blunders, such as the inappropriate use of analysis of covariance and the use of subjects as experimental units when intact classes have been assigned to treatments. More disturbing still is the investigators' apparent ignorance that statistical assumptions are being violated.

Gorman (1968) analyzed 293 articles and dissertations on problem solving in terms of their experimental design as defined by Campbell and Stanley (1963). Of the studies analyzed, 178 were removed from consideration for not actually being studies of problem solving in mathematics, for not using as subjects students in grades 1-6, for not being published in the years from 1925 to 1965, for lacking internal validity, or for not being available for examination. The remaining 115 studies were closely examined for internal validity. Finally, the number of studies which met criteria was 37. The recommendations from the accepted studies concerning the teaching of problem solving in grades 1-6 were centered in the following areas.

a. Effect of using the following methods:

(1) Systematic instruction[1]

(a) Systematic instruction in which students are asked to explain how a problem is to be solved and why a particular process is appropriate produces greater gains in problem solving than mere presentation of many problems.

(b) The development of understanding is a gradual process that is aided by systematic instruction in the four fundamental processes.

(c) The development of understandings of the four fundamental processes is a vital factor in the improvement of problem solving.

(2) Intensive study of vocabulary[2]

(a) Pupils who studied quantitative vocabulary using the direct study techniques (enable the child to establish a three-way association between the written symbol, the sound of the term, and at least one of its meanings) achieved significantly higher on a test of arithmetic problem solving and concepts than pupils who had not devoted special attention to the study of quantitative vocabulary.

(b) The direct study of quantitative vocabulary does not tend to result in an improvement in general vocabulary or in reading comprehension.

(c) The direct study of quantitative vocabulary is not more effective with one sex than with the other.

(d) The direct study of quantitative vocabulary is a method that is more effective with pupils of above average or average intelligence than it is with pupils of below average intelligence.

---

[1] Angela Pace, "The Effect of Understanding on the Reorganization and Permanence of Learning" (unpublished Doctor's dissertation, Syracuse University, 1959).

[2] Louis Frederick VanderLinde, "An Experimental Study of the Effect of the Direct Study of Quantitative Vocabulary on Arithmetic Problem Solving Ability of Fifth Grade Pupils" (unpublished Doctor's dissertation, Michigan State University, 1962).

(3) Estimating answers[1]

Practice in estimating answers to arithmetic problems is of no more value to sixth grade pupils than is the traditional practice in the solution of such problems.

(4) Group experience[2]

Despite the superiority of the work performed by groups as compared to individual efforts in situations involving written problem solving, there is no significant improvement in the ability of subjects to solve problems when trained in groups as compared to subjects who have worked by themselves continuously when evaluated under circumstances in which each subject must rely upon his own resources.

(5) Cuisenaire materials[3]

Use of Cuisenaire materials in an elementary mathematics program resulted in significantly less achievement in computation and reasoning than was evident when such materials were not used.

b. Comparison of methods:

(1) Cooperative versus individual effort[4]

(a) Children working together in pairs do solve more problems correctly than each child could do working alone.

(b) Children working together in pairs do require more time to solve problems than each child would do working alone.

---

[1] John W. Dickey, "The Value of Estimating Answers to Arithmetic Problems and Examples," The Elementary School Journal, XXV (September, 1934), 24-31.

[2] Bryce Byrne Hudgins, "The Effects of Initial Group Experience upon Subsequent Individual Ability to Solve Arithmetic Problems" (unpublished Doctor's dissertation, Washington University, 1958).

[3] Robert Albert Passy, "How Do Cuisenaire Materials in a Modified Elementary Mathematics Program Affect the Mathematical Reasoning and Computational Skill of Third Grade Children?" (unpublished Doctor's dissertation, New York University, 1963).

[4] Samuel F. Klugman, "Cooperative Versus Individual Efficiency in Problem Solving," Journal of Educational Psychology, XXXV (February, 1944), 91-99.

(2) Drill method with insight method[1]

(a) If skill in computation and solving verbal problems is the chief goal of instruction, the method a teacher employs should be determined by her own predilections.

(b) If more generalized outcomes of instruction, particularly the ability to think mathematically, are significant goals, it makes a difference how pupils are taught.

(c) Pupils of relatively low ability and good achievement learn better under a drill method.

(d) Pupils of relatively high ability and low achievement learn better under a meaning method.

(3) Association, analysis and vocabulary[2]

The association method, or that technique by which difficult or incorrect problems are associated with a model, produces greater gain in student performance in problem solving than the analysis or vocabulary methods. (The analysis method refers to the step-by-step approach to problem solving while the vocabulary method involved the completion of mathematical problems with the correct term.)

(4) Dependencies, conventional-formula, and individual[3]

(a) The conventional-formula method of problem solving (i.e., four steps: asked, given, how, answer) provided the least gain in ability when compared with the individual (absence of any formula) or dependencies method (graphic or diagrammatical).

---

[1] G. Lester Anderson, "Quantitative Thinking as Developed Under Connectionists and Field Theories of Learning," Learning Theory in School Situations (University of Minnesota Studies in Education, No. 2; Minneapolis: University of Minnesota Press, 1949).

[2] C. L. Thiele, "A Comparison of Three Instructional Methods in Problem Solving," Research on the Foundations of American Education, Official Report of the American Educational Research Association (Washington, D. C.: American Educational Research Association, 1939).

[3] Paul R. Hanna, "Arithmetic Problem Solving: A Study of the Effectiveness of Three Methods of Problem Solving" (unpublished Doctor's dissertation, Teacher's College, Columbia University, 1929).

(b) When fourth and seventh graders are considered as a whole, there is no difference between the results of the dependencies and individual methods.

(c) When the work of fourth graders alone is analyzed, the data indicate that the dependencies method is superior to the individual approach.

(5) Formal analysis and graphic analysis[1]

Neither the conventional (formal analysis) nor the dependencies method (graphic analysis) produced changes that were statistically significant with respect to the following:

(a) the grade level on which they were used;
(b) the ability levels within the grade (average, superior, inferior);
(c) retention of ability in problem solving.

(6) Action Sequence, Wanted-Given, and Practice-Only[2]

(Action Sequence refers to a program focusing on what is going on, what events went on, what is being done, what is done, what was the sequence of events, etc., in situations from which the meaning or "attributes" of an operation is to be abstracted. This program emphasizes what one does mentally or physically when one is adding, subtracting, etc. Hence, the operations are conceived of in terms of their characteristic action-sequence. In other words, the operations are relationships, or have structures, the relational attributes of which are action-sequences.

Wanted-Given refers to the program which focuses on the goals and "tools," the "why" and "with what," the ends and means, the wanted and givens in situations from which the meaning or "attributes" of an operation is to be abstracted. This program emphasizes "why" and "with what" one adds, subtracts, etc. Hence, the operations are conceived of in terms of their characteristic ends-means. Or, in other words, the operations are relationships, or

---

[1] Ralph D. Horsman, "A Comparison of Methods of Teaching Verbal Problems in Arithmetic in Grades Five, Six, Seven, and Eight" (unpublished Doctor's dissertation, University of Pittsburgh, 1940).

[2] John Warner Wilson, "The Role of Structure in Verbal Problem-Solving in Arithmetic: An Analytical and Experimental Comparison of Three Problem-Solving Programs" (unpublished Doctor's dissertation, Syracuse University, 1964).

> have structures, the relational attributes of which are wanted-givens.
>
> Practice-Only is a non-specified structures program which provides no direct instruction concerning the problem situation.)
>
> The major difference in the three programs is in the contents of the thought process of the child as he analyzes a problem and chooses the operation.
>
> (a) Emphasizing those attributes of the arithmetic operations termed the Wanted-Given produces statistically significant improvement in verbal problem solving ability in arithmetic.
>
> (b) Emphasizing the Wanted-Given attributes of the operations produces statistically significant improvement in verbal problem solving ability than comparable emphasis of the Action-Sequence attributes of the operations.
>
> (c) Emphasis on the Wanted-Given attributes produces a greater statistical improvement in verbal problem solving than the mere provision of practice.
>
> (d) Emphasizing the Action-Sequence attributes of the operations produces no statistically significant improvement in verbal problem solving.

We will give more consideration to Wilson's findings on following pages. First, however, the summaries of the journal-published aritcles on problem solving in arithmetic, grades 1-8, for the years 1900-1968 by Suydam and Riedesel (1969) are presented. In all, Suydam, Riedesel and staff examined all the articles on arithmetic published in 47 American and English journals and synthesized the findings of 1,104 of the studies which were reports of actual experiments. Dissertations were also listed, but not synthesized. The findings from research carried out in the period 1950-1968 were emphasized, but significant findings from before the advent of modern mathematics, 1950, were also included. A partial list of their "Answers from Research: Problem Solving" follows.

How do pupils think in problem solving?

Studies by Stevenson (1925) and Corle[1] (1958) reveal that pupils often give little attention to the actual problems; instead, they almost randomly manipulate numbers. The use of techniques such as 'problems without number' can often prevent such random attempts.

What is the importance of the problem setting?

Researchers such as Bowman (1929, 1932), Brownell[2] (1931), Hensell (1956), Evans (1940), Sutherland (1941), Wheat (1929), and Lyda and Church[1] (1964) have explored the problem setting. Findings are mixed, with some researchers suggesting true-to-life settings while others suggest more imaginative settings. While the evidence appears to be unclear, one thing does emerge: problems of interest to pupils promote greater achievement in problem solving. With today's rapidly changing world it seems unreasonable that verbal problems used in elementary school mathematics could sample all of the situations that will be important to pupils now and in adult life. Perhaps the best suggestion for developing problem settings is to take situations that are relevant for the child. Thus, a problem on space travel may be more 'real' to a sixth grader than a problem based upon the school lunch program.

How does the order of the presentation of the process and numerical data affect the difficulty of multi-step problems?

Burns and Yonally[2] (1964) found that pupils made significantly higher scores on the test portions in which the numerical data were in proper solution order. Berglund-Gray and Young[1] (1932) found that when the direction operations (addition and multiplication) were used first in multi-step problems, the problems were easier than when inverse operations (subtraction and division) were used first. Thus, an 'add-then-subtract' problem was easier than a 'subtract-then-add' problem.

What is the effect of vocabulary and reading on problem solving?

Direct teaching of reading skills and vocabulary directly related to problem solving improves achievement (Robertson, 1931; Dresker,[2] 1934; Johnson, 1944;[1] Treacy, 1944; VanderLinde,[2] 1964).

How does wording affect problem difficulty?

Williams and McCreight[2] (1965) report that pupils achieve slightly better when the question is asked first in a problem. Thus, since the majority of textbook series place the question last, it is suggested that the teacher develop and use some word problems in which the question is presented first.

---

[1] A study rejected by Gorman.

[2] A study accepted by Gorman.

What is the readability of verbal problems in textbooks and in experimental materials?

Heddens and Smith (1964) and Smith and Heddens (1964) found that experimental materials were at a higher reading difficulty level than commercial textbook materials. However, they were both at a higher level of reading difficulty than that prescribed by reading formula analysis.

What is the place of understanding and problem solving?

Pace[1] (1961) found that groups having systematic discussion concerning the meaning of problems made significant gains. Irish[1] (1964) reports that children's problem solving ability can be improved by (1) developing ability to generalize the meanings of the number operations and the relationships among these operations, and (2) developing an ability to formulate original statements to express these generalizations as they are attained.

Should the answers to verbal problems be labeled?

While Ullrich (1955) found that teachers prefer labeling there are many cases in which labeling may be incorrect mathematically. For example:

Incorrect	Correct
10 apples	10
+6 apples	+6
16 apples	16 apples

Does cooperative group problem solving produce better achievement than individual problem solving?

Klugman[2] (1944) found that two children working together solved more problems correctly than pupils working individually. However, they took a greater deal of time to accomplish the problem solutions. Hudgins[2] (1960) reported that group solutions are no better than the independent solutions made by the most able member of the groups.

What is the role of formal analysis in improving problem solving?

The use of some step-by-step procedures for analyzing problems has had wide appeal in the teaching of elementary school mathematics. Evidence by Stevens (1932), Mitchell[1] (1932), Hanna (1930), Bruch (1953), and Chase (1961) indicated that informal procedures are superior to following rigid steps such as the following: 'Answer each of these questions: (1) What is given? (2) What is to be found? (3) What is to be done? (4) What is a close estimate of

> the answer? and (5) What is the answer to the problem?' If this analysis method is used, it is recommended that only one or two of the steps be tried with any one problem.
>
> What techniques are helpful in improving pupils' problem solving ability?
>
> Studies by Wilson[1] (1922), Stevenson[1] (1924), Washburne (1926), Thiele[2] (1939), Luchins[1] (1942), Bemis and Trow (1942), Hall[1] (1942), Klausmeier (1964), and Riedessel[1] (1964) suggest that a number of specific techniques will aid in improving pupils' problem-solving ability. These techniques include: (1) using drawings and diagrams, (2) following and discussing a model problem, (3) having pupils write their own problems and solve each others' problems, (4) using problems without numbers, (5) using orally presented problems, (6) emphasizing vocabulary, (7) writing mathematical sentences, (8) using problems of proper difficulty level, (9) helping pupils to correct problems, (10) praising pupil progress, and (11) sequencing problem sets from easy to hard.

A later summary of research on problem solving and recommendations for teaching problem solving was reported by Riedesel (1969, pp. 54-55). A partial list of his suggestions follows.

> 1. While the improvement of computation is important to problem-solving ability, the improvement of computation alone has little, if any, measurable effect upon reasoning and problem solving.
>
> 2. Motivation is essential to effective learning. To assure optimal achievement, pupils must be interested in the problem-solving program. It has been found that pupils react well to a variety of problem settings. The use of 'real-life settings' may be helpful. Also, it has been found that use of problems from old U.S. and new foreign textbooks increases pupil interest.
>
> 3. Children at all levels of problem-solving ability are receptive to supplemental 'puzzle-type' or enrichment problems.
>
> 4. The use of formal analysis--that is requiring pupils to answer each of the following questions: (1) What is given? (2) What is to be found? (3) What operation do you use? (4) What is an estimate of the answer? (5) What is the answer to the problem?--does not produce superior results in problem solving. However, the use of one of the steps as a focus for a lesson does improve problem solving. For example, on a given day the task given to the pupils might be this: 'Make an estimate of an answer to the problem. You do not have to compute your answer.'

5. The following teacher practices improve problem solving:
   a) Provide problems of appropriate difficulty level.
   b) Guide pupils to use a method for getting started.
   c) Aid pupils in the analysis of information.
   d) Give pupils encouragement to proceed, and praise them when they perform some process correctly.
   e) Aid pupils in verifying final solutions.
   f) Start pupils with easy problems which they most certainly can get correct with a reasonable amount of effort.

6. The problem-solving program should be started early. As soon as a child begins to work with sets, he can begin to solve orally presented verbal problems. Thus an important part of the kindergarten program should be verbal problem solving.

7. Problems may be used at various times in a unit: at the beginning to introduce a topic, as a unit progresses, and as review and maintenance.

8. A variety of computational types should be presented in most problem-solving lessons. When pupils find that all problems for a day's lesson involve one operation, the task is actually only one of computational practice.

9. Tape recordings of the textbook problems can be used with pupils who have difficulty reading problems.

10. A best technique has not been found for problem solving. However, the following techniques have been found to increase problem-solving ability:
   a) Make use of mathematical sentences in solving single and multi-step problems.
   b) Make use of drawings and diagrams as a technique to help pupils solve problems.
   c) Make use of orally presented problems.

In all the studies and reviews of studies examined thus far, there is general agreement that students can be taught to solve problems. However, as can be seen, there is disagreement on how it should be done and of which reports should be accepted as representing good research. Formal analysis, where five or more steps are used to solve a problem, has not been shown to produce significant gains in problem solving. However when one or two of the steps are emphasized, significant gains have resulted. Wilson's study (1967) compared the wanted-given approach to teaching problem solving with the action-sequence approach. In the wanted-given approach the child was to (Wilson, 1967, p. 488):

> 1. Recognize the wanted-given structure of the problem.
> 2. Express this structure as an equation.
> 3. Compute by using the operation indicated by the equation.

In the action-sequence program the child was to (p. 487):

> 1. 'See' or recognize the real or imagined action-sequence structure of a problem.
> 2. Express the action sequence in an equation.
> 3. Compute, using the operation indicated if the equation is direct; or if the equation is indirect, imagine an appropriate second action sequence, express it as an equation, and compute using the operation indicated.
> 4. Check by rewriting the equation with the answer in the proper position.

The findings of the wilson study were cited earlier by Gorman (1967, p. 94) except that in his quote from the Wilson article (Wilson, 1967, p. 495) Gorman did not include the qualification "For the types of one-step verbal problems used in this study and for the age and grade level of the children involved--" which preceded the four main conclusions. The point is that Wilson's study used only one-step problems.

In Wilson's treatment (1967, p. 55), the wanted-given structure of the problem was seen as a means-ends relationship. Wilson assumed that the child "saw" the structure of the problem and expressed the relationship in equation form (p. 57). The equations were always to be written by the student in direct rather than indirect form. That is, $12 - 5 = n$ rather than $n + 5 = 12$. A sample problem from each category used by Wilson in his study best illustrates his treatment.

Sample 1. Problem (p. 60)

Bob had 9 marbles. Dick gave him 3 marbles.
How many marbles did Bob have then?

Classification

A problem in which the size of a total is wanted and the sizes of its parts are given.

Wanted-given structure of addition

$9 + 3 = n$ or $3 + 9 = n$

Sample 2. Problem (p. 61)

Bob had 9 marbles. After Dick gave him some more marbles, he had 12 marbles. How many did Dick give Bob?

Classification

A problem in which the size of one part is wanted and the sizes of the total and the other part are given.

Wanted-given structure of subtraction

$12 - 9 = n$

Sample 3. Problem (p. 61)

Bob had 12 marbles. He lost 3 of them.
How many marbles did Bob have then?

Classification

A problem in which the size of one part is wanted and the sizes of the total and the other part are given.

Wanted-given structure of subtraction

$12 - 3 = n$

Sample 4. Problem (p. 62)

Bob had some marbles. He lost 3 of them. Then he had 9 marbles left. How many marbles did Bob have to begin with?

Classification

A problem in which the size of a total is wanted and the sizes of its parts are given.

Wanted-given structure of subtraction

$9 + 3 = n$ or $3 + 9 = n$

Unfortunately, the proportion of problems requiring each of the four basic operations in Wilson's texts was not the same. Of the 30 items on each of his posttest and retention tests, 9 were addition problems, 13 were subtraction problems, 3 were multiplication problems, and 5 were division problems. As indicated by the examples, the exercises were quite easy and few purely computational errors were made. Wilson (1967) concluded that selecting the proper operation was of primary importance. "The group which became superior in choosing correct operations also became superior in obtaining the correct answers (p. 229)."

Marilyn Zweng (1968) criticized the action-sequence program used by Wilson (1967) for not being flexible enough in its instructional approach. That is, although there are several ways to solve most problems, she contended that the children in the action-sequence group were led to believe there was only one correct mathematical model for each problem. Even though Wilson's article was necessarily a shortened version of his dissertation and did not include all the details, there was not enough evidence to support Zweng's criticism in full. However, the extent of control required to maintain the distinctiveness of each treatment in Wilson's study did tend to restrict the flexibility more than one might expect in each treatment had it been carried out separately under less rigorous conditions. Each treatment required the student to recognize a number of situations in which the structure was different or the action sequence varied. In the action-sequence treatment, the child visualized or imagined the action taking place in the problem and was required to write the corresponding mathematical sentence in exactly the order indicated by the action sequence (Wilson, 1967, p. 26). This required an almost exact translation of the action taking place into mathematical terms as expressed by an equation. The wanted-given treatment, on the other hand, required the student to recognize the structure of the problem in terms of a direct mathematical sentence (Wilson, p. 43). In some cases expressing the relationship in a direct equation amounted to two steps for students in the action-sequence treatment. They were required to imagine a second step, the direct equation, if the action sequence indicated an indirect equation. That is, if the action sequence suggested an equation like $10 + n = 23$, the student had to imagine some action sequence that would have a direct form such as $23 - 10 = n$. The student then had to perform the operation indicated in the second step to find n.

The modification of Wilson's wanted-given treatment used in the present study was aimed at reducing the number of cases a student had to keep in mind while trying to solve a problem and at the same time make the solution process more meaningful. As suggested earlier by Thiel (1939) and Mitchell (1932), analysis appears most efficient when only one or two steps are required. For this reason the wanted-given approach used in the present study emphasized the notion of two alternatives, that is, two steps. First, students were instructed in the use and meaning of the terms "sum," "addend," "product," and "factor." Following this, they were given practice in solving simple equations, such as $15 + 10 = n$. Next, students were taught that to find the value of n in an equation such as $15 + n = 25$ they had to solve the equation $n = 25 - 15$.

In a similar manner, students were taught that to solve for n in $25 \times n = 20$ they had to solve $n = 20 \div 5$.

The instructional program emphasized that there were basically just two kinds of problems, sum problems and product problems. A sum problem is one in which two or more quantities are being combined. Once a student established that the problem was a sum-type problem, only two operations were possible, addition or subtraction. If the problem involved direct addition, such as $a + b = c$, or if it had one of the addends missing (wanted) in the form $a + n = c$, the solution was $n = c - a$. Similarly for product problems, once the student determined it was a product-type problem, just two operations were possible, multiplication or division. The easier of the two operations was multiplication, multiplying the factors $a \times b = c$, to find the product. If one factor was missing (wanted) and the equation was the form $a \times n = c$, then the solution was $n = c \div a$.

Beginning with one-step problems, the program led the student to make choices between two alternatives at each step. Keeping the number of alternatives to two is in keeping with the findings of earlier studies (Thiel, 1939; Mitchell, 1932) and facilitates both understanding and instruction through emphasis on vocabulary (Vanderlind, 1964).

Figure 1 presents the decision structure of the program. At level 1, the student must decide whether the problem is a sum- or product-type problem. At level 2, he again must make a choice of whether to supply a sum, an addend, a product, or a factor.

The transfer of this approach to multiple-step problems is direct. The decision structure is simply a nested set of wanted-given sequences in which the rules from case to case do not vary. This sequence is presented graphically in Figure 2. For example, a factor may itself be a sum. The student must find the sum (step 1) and then use the sum as one

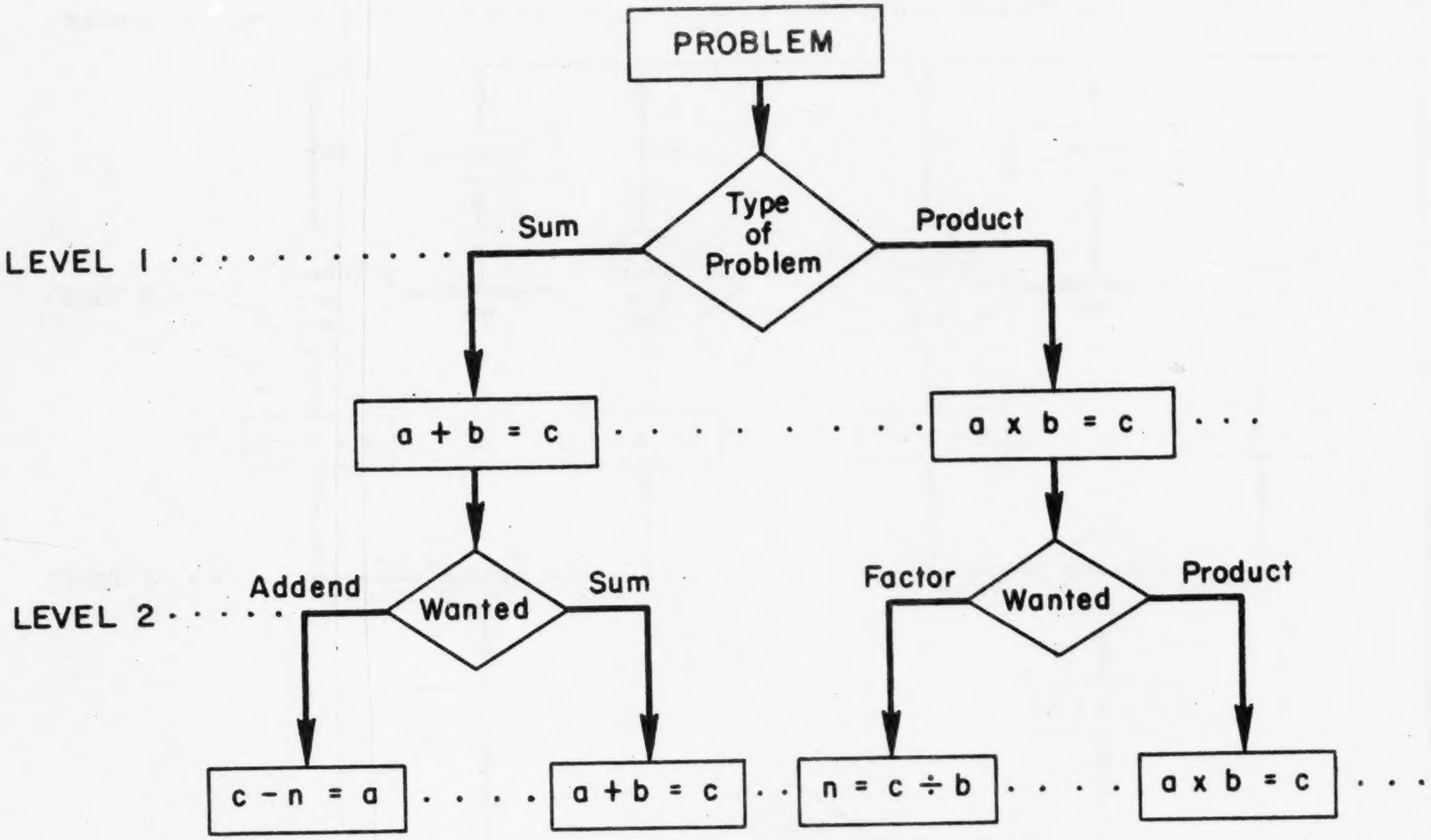


Figure 1. Decision structure of the Modified Wanted-Given Program for simple one-step verbal problem solving indicating decision levels.

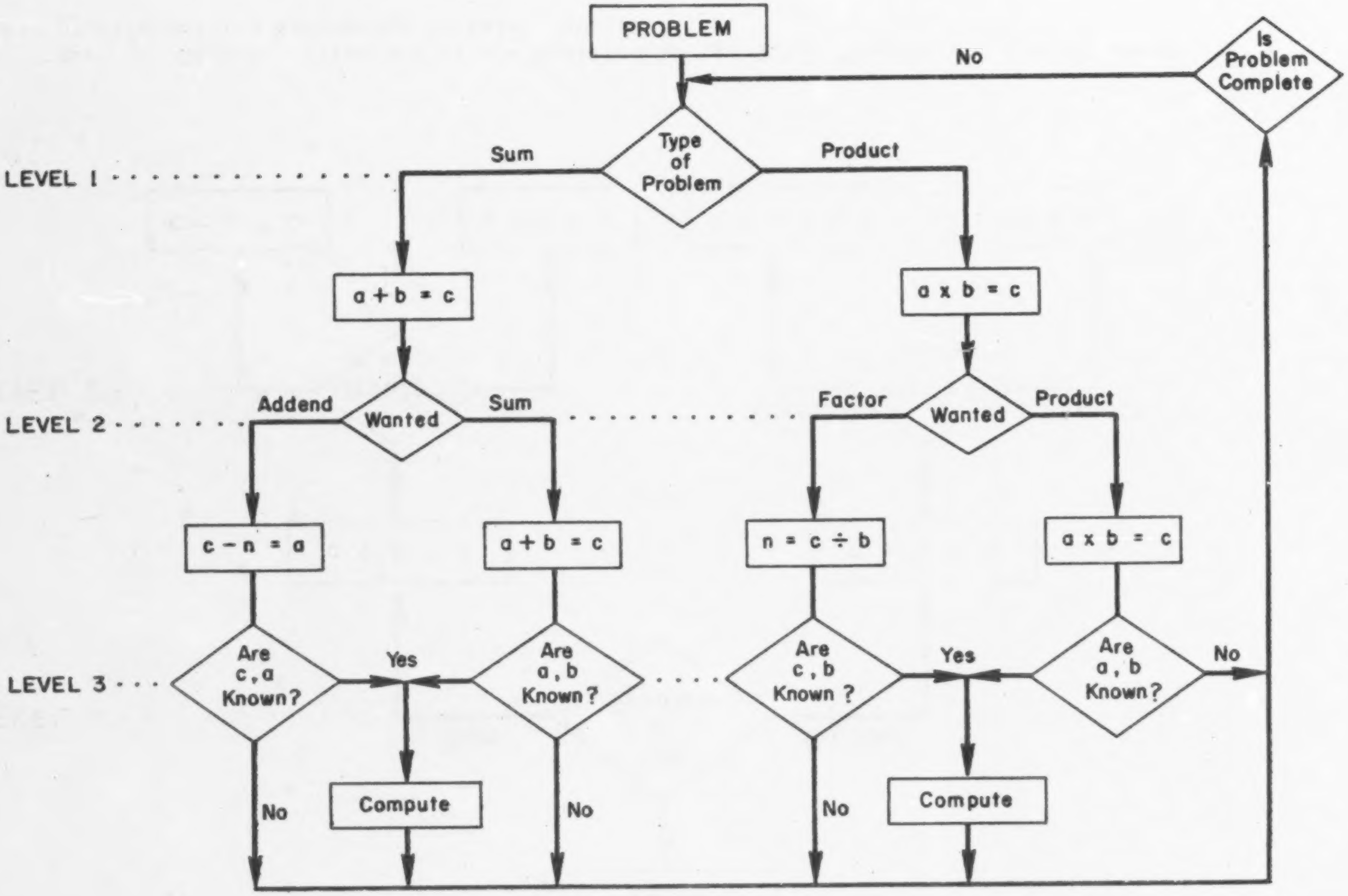


Figure 2. Decision structure of the Modified Wanted-Given Program for multi-step verbal problem solving.

of the factors to find the product (step 2). It is this second step that is often forgotten by fifth graders. They are, in the experience of the writer, inclined to perform one step and feel that they have solved the problem since they have "done something." It was assumed that teaching students to first recognize the overall structure of the problem would improve their understanding of the problem and reduce the number of occurrences of only partially completed problems.

By way of comparison, the first example used to illustrate Wilson's approach (p. 60) was almost identical in structure to that used by a student in the present program. Rather than being classified as a problem in which the size of the total was wanted, the problem was classified by the present program as a sum problem. What was wanted was the sum. The addends were given. Therefore the operation was addition, $9 + 3 = 12$ or $3 + 9 = 12$.

The second example Wilson gave (p. 61) would also be classified as a sum problem of the form $9 + n = 12$. However, since one of the addends was missing, the problem had to be solved by using the equation $n = 12 - 9$.

The third case cited by Wilson (p. 61) was also a sum problem in terms of the present program. The equation was $n + 3 = 12$. The equation $n = 12 - 3$ had to be used to find the number of marbles remaining.

The fourth case was yet another example of a sum problem. The addends were given. The student was asked to find the sum.

Several pages from the instructional treatment which illustrate various stages of the program, are included in Appendix A.

In summary, the key differences between Wilson's program and the approach used in this study are:

1. Each problem, as a first step, was identified according to one or two basic types, sum- or product-type problem, rather than the wanted-given, part-whole relationship used by Wilson (1967, p. 43).
2. Wanted-given was taught in terms of sum, addends, factor, product, rather than in terms of parts and whole (Wilson, p. 36-38).
3. The number of rules to be learned and remembered by the student was smaller as was the number of different types of problems (see Figure 1).
4. The meaning of the terms "sum," "addends," "product," and "factor" were stressed throughout the program. These terms were not emphasized in Wilson's study.
5. Where appropriate, students were encouraged to write indirect equations for the first step as expressions of the problem type rather than required direct equations in each case (Wilson, p. 43). Direct equations were treated as the natural solution to indirect equations. That is, $12 - 9 = n$ is the solution to $9 + n = 12$.

The instructional program itself was presented to students in programmed-booklet form, one booklet each day for 16 days. The program was similar to that of The Productive Thinking Program in approach, in that students were to follow a story line about a boy, Bill, and a man, Mr. Smith, as they encountered and solved a series of mathematical problems. The man in the story was trying to help the boy become a good problem solver and encouraged the boy, in turn, to help the readers to become good problem solvers as well. The linear program provided for immediate feedback, cues, and step-by-step instruction using small steps.

SUMMARY AND HYPOTHESES

Researchers, such as Olton, Crutchfield, and Covington, have argued for the transferrability of the problem-solving skills developed through systematic instruction in _The Productive Thinking Program_. The results of several studies confirm their arguments in some cases, but not in others. Failure to produce significantly greater achievement levels in mathematics as transfer from training in _The Productive Thinking Program_ in studies to date may have been due to the fact that the mathematical tasks used in some of these studies were too simple to demonstrate the full effects of the training. The lack of sex differences in achievement for students using the programmed materials may be an indication of the teacher independence of the treatment.

On the other hand, there are only general suggestions from previous research on how one ought to go about teaching students to solve problems. Any educator who wishes to prepare an instructional program in problem solving must glean, from the best of what others have attempted, the techniques he feels most appropriate in light of previous research and theory.

The purpose of this study was to test the hypotheses that:

1. There will be no significant differences in scores on posttest measures of problem solving skills between students who receive training in general problem solving techniques as presented in _The Productive Thinking_

Program and students who are taught more specific skills for solving problems in mathematics in a totally mathematics context using the Modified Wanted-Given Program.

2. Students in either of the treatment groups, The Productive Thinking Program or the Modified Wanted-Given Program, will not achieve significantly higher scores on measures of problem skills than students in control classes who receive no special instruction in problem solving.
3. No significant differences will be apparent between the achievement of boys and girls on tests of problem solving skills in the treatment groups.

We turn now to the design of the experiment itself.

CHAPTER II

EXPERIMENTAL DESIGN

As stated in Chapter I, the purpose of this study was to compare experimentally the differential effects of three instructional programs on problem solving and on measures of problem-solving skills of fifth-grade students.

PROCEDURE

Three treatments--The Productive Thinking Program, the Modified Wanted-Given Program, and The Control (no treatment)--were administered to eight classes of fifth-grade students in the Cupertino Union School District near San Jose, California. Six of the classes received the experimental treatments, on a random basis, and two classes served as controls. The three treatments and two sex variables constituted the 3 × 2 factorial design of this study, which was as follows:

$$\begin{array}{cccc} O_1 & X_1 & O_2 & O_3 \\ O_1 & X_2 & O_2 & O_3 \\ O_1 & & O_2 & O_3 \end{array}$$

where O_1 is the pretest, X_1 The Productive Thinking Program treatment, X_2 the Modified Wanted-Given Program treatment, O_2 the posttest, and O_3 the follow-up test which was given seven weeks after the posttest. The schedule for each event was as follows:

Pretest (in all four schools) Thursday, October 22, 1970.

Treatment (in three schools) Monday, October 26-Tuesday,

November 16, 1970 (there was one school holiday in
the treatment period).

Posttest (in all four schools) Wednesday, November 18, 1970.

Follow-up test (in all four schools) Thursday or Friday,
January 7 or 8, 1971.

Subjects. The 261 subjects of this study were all students in fifth-grade classes in four public schools in the Cupertino Union School District in Santa Clara County, California. The distribution of subjects by school, class, treatment, and sex is shown in Table 1. All students (a total of 287 that included 26 special students identified and classified by the school as either gifted or educationally handicapped to avoid discrimination of any sort) took part in the study. The data from the tests of the special students were not included in the analysis of the results.

Assignment to treatment. Both instructional treatments were presented to students in programmed booklets. In each treatment one booklet was given to each student each day for 16 consecutive school days. Booklets were randomized using a table of random numbers (Fisher and Yates, 1957) and handed out daily to students in the three experimental schools within each class beginning with the first day of the treatment period following the pretest. Each student remained in the program to which he was assigned on the first day for the remainder of the experimental period. The control classes received no treatment.

Administration of the treatments. As noted by Olton (1969) and others earlier, permitting the teacher to discuss the treatment with the class had a significant influence on student achievement. To control the teacher variable, aides were hired to administer the tests and daily

TABLE 1

Distribution of Students by Class, Sex, and Treatment

Sex	Treatment Classes												Control Classes		Total
	A		B		C		D		E		F		G	H	
	P[a]	WG[b]	P	WG	P	WG	P	WG	P	WG	P	WG			
Boys	7	7	10	7	4	5	7	6	8	7	7	10	20	22	127
Girls	11	8	6	12	11	10	4	5	8	12	10	9	16	12	134
Subtotal	18	15	16	19	15	15	11	11	16	19	17	19			
Total	33		35		30		22		35		36		36	34	261

[a] The number of students in The Productive Thinking Program.
[b] The number of students in the Modified Wanted-Given Program.

treatments in each class. Of the three aides, one was the wife of a graduate student, one had prior experience dealing with children in a school situation as a substitute teacher for a short time, and the other had served as a teacher aide in a junior high school the previous year. All aides were given the pretests before the experiment began to familiarize them with the procedure. The purpose of the experiment and the role they were to play were carefully explained. Each aide was given a set of printed instructions to follow the first few days of the experiment (see Appendix B). Further, aides were instructed not to discuss the problems with students, although they could answer any questions and act supportively and pleasantly. The aides were in charge of the class during that period of the day when the treatments or tests were given. All treatments and tests were given before noon each day in each school.

The aides distributed the instructional materials at the beginning of each class period. At the close of each class period the instructional booklets were picked up by the aides so that no booklets remained at the school. The classroom teachers did not participate in the administration of the treatments and generally used the time to catch up on other work; in fact the classroom teachers did not even see a set of the instructional materials. Their only contact with the experiment was during the initial meeting held at each school at which time the purpose of the experiment was explained and their cooperation was requested.

The daily treatments were given during the regular mathematics period, and the entire period was used each day. The only mathematics instruction the treatment classes received during the period of the experiment was that contained in the instructional treatments. Those students who were assigned to *The Productive Thinking Program* received no mathematics instruction for 16 days, because that program contained none. Those students who were assigned to the Modified Wanted-Given Program received instruction emphasizing the solution of problems, as described earlier, rather than computational skills. The control classes followed the state-adopted text which contained no special emphasis on problem solving and few problem-solving exercises. The teachers of the control classes agreed not to instruct students in problem solving during the time the experiment was in progress.

Development of treatments. *The Productive Thinking Program* (Covington, Crutchfield, and Davies, 1966) is a copyrighted, commercially available program. One hundred sets of the instructional materials were purchased for use in the study. The program itself is described in detail in Chapter I.

The Modified Wanted-Given Program was developed by the author for this study. Its content and approach are also described in Chapter I. Preliminary versions of the daily lessons were prepared and tested in a single fifth-grade class which was in a school in the same general socioeconomic area and school district as the classes that participated in the experiment. In all, 10 students worked through the first 10 lessons of the program. Based on their performance and the teacher's recommendations, the program was revised. The revised version in 16 daily lessons constituted the Modified Wanted-Given Program treatment. Since the first 10 lessons contained the instructional sequence, piloting only the first 10 lessons was considered sufficient. Lessons 11-16 contained internal tests and extensive reviews of the material introduced in the first 10 lessons.

Measuring instruments and scoring procedures. The pretest consisted of four scales--_Figure Classification_, _Working with Numbers_, _Arithmetic Reasoning_, and _Hidden Figures_--which were to be covariates with the posttest and follow-up test scales. Statistics on each are given here, along with a brief description of the skill it was intended to measure. A copy of the pretest is given in Appendix C.

The _Figure Classification Test_ is designed to measure a student's ability to discover rules that explain things. This scale is an adaptation of the Figure Classification Test developed by the Educational Testing Service (ETS) (1962), which itself is an adaptation of a University of North Carolina version of Thurstone's test of the same name. The ETS form of the test consisted of two parts of 14 items each with a time limit of 8 minutes. Many of the items were thought to be too difficult for fifth-grade students.

A form of the Figure Classification Test was pilot tested in four fifth-grade classes with a total of approximately 120 students. These four fifth-grade classes did not participate in the later study. The nine items which correlated most highly with the total test score were selected for use in the present study.

The instruction for the Figure Classification Test, together with examples, are given on the following page. No scale statistics were available from ETS for this test. The time limit for the revised form, nine items, was set at 10 minutes to decrease the importance of speed as a factor.

The second scale, Working with Numbers, was a test used in the five-year National Longitudinal Study of Mathematical Abilities (NLSMA) conducted by the School Mathematics Study Group at Stanford University. A description of the scale, the scale statistics, and a sample test item follow (NLSMA Reports, No. 4, 1968, p. 33).

Spring Year 2
Grade 5

X307 WORKING WITH NUMBERS (12 items; 20 minutes) This scale is designed to measure ability to perform operations using whole numbers according to written directions. It has five items in common with X313.

EXAMPLE: The sum of the odd numbers less than 4 and the even numbers less than 9 is
(A) 11 (B) 13 (C) 24 (D) 42 (E) 45

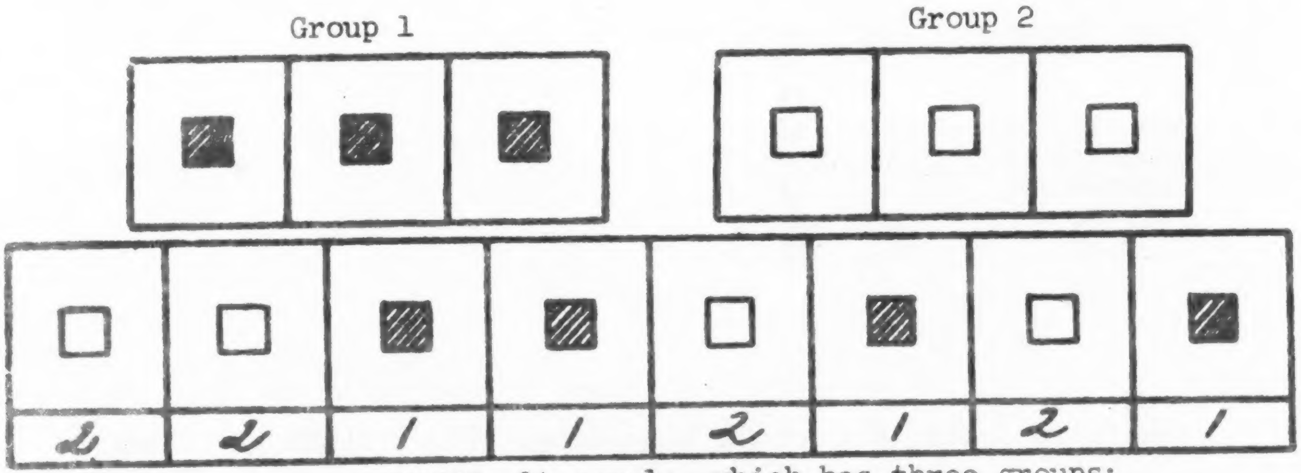
SCALE STATISTICS:

MEAN = 5.66 ALPHA = 0.68 SAMPLE SIZE= 1332
ST.DEV= 2.62 ERR.MEAS= 1.49

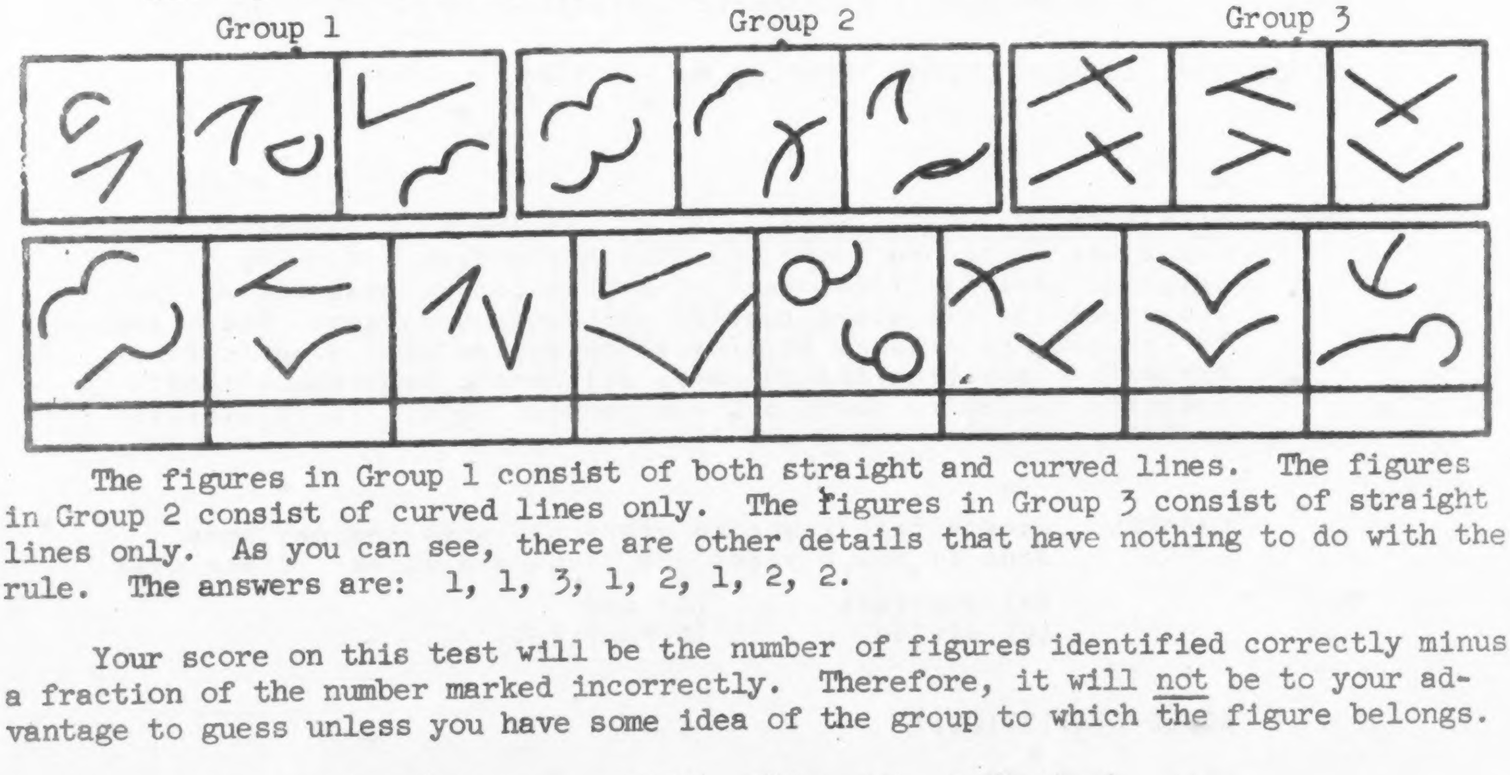
FIGURE CLASSIFICATION - I-3

This is a test of your ability to discover rules that explain things. In each problem on this test there are either two or three groups, each consisting of three figures. You are to look for something that is the same about the three figures in any one group and for things that make the groups different from one another.

Now look at the sample problem below. In the first line, the figures are divided into Group 1 and Group 2. The squares in Group 1 are shaded and the squares in Group 2 are not shaded. In the second line a 1 has been written under each figure that has a shaded square as in Group 1. A 2 has been written under each figure with an unshaded square as in Group 2.



Now try this more difficult sample, which has three groups:



The figures in Group 1 consist of both straight and curved lines. The figures in Group 2 consist of curved lines only. The figures in Group 3 consist of straight lines only. As you can see, there are other details that have nothing to do with the rule. The answers are: 1, 1, 3, 1, 2, 1, 2, 2.

Your score on this test will be the number of figures identified correctly minus a fraction of the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you have some idea of the group to which the figure belongs.

You will have 10 minutes to complete this part of the test.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Copyright © 1962 by Educational Testing Service. All rights reserved.
A University of North Carolina Adaptation of a test by L. L. Thurstone

ITEM STATISTICS:

ITEM INDEX	1	2	3	4	5	6	7	8	9	10
P	69	79	52	45	74	42	51	37	31	17
ADJ. P	69	79	52	45	74	42	51	37	31	17
BISERIAL	56	52	56	39	47	53	49	48	37	01
PERCENT NT	0	0	0	0	0	0	0	0	0	0
PAGE NO.	185	185	185	185	186	186	186	187	187	187
ITEM NO.	1	2	3	4	5	6	7	8	9	10

ITEM INDEX	11	12
P	31	38
ADJ. P	31	38
BISERIAL	39	9
PERCENT NT	0	2
PAGE NO.	188	188
ITEM NO.	11	12

The third scale, Arithmetic Reasoning, also a NLSMA test, was adapted from the Necessary Arithmetic Operations Test. The description of the test follows (NLSMA Reports, No. 4, 1968, p. 180).

PX217 NECESSARY ARITHMETIC OPERATIONS I (15 items; 5 minutes) This scale was patterned after the French Kit Form R-4. The original test has items requiring both one and two operations. The items in this scale require only one operation. The scale is intended to measure ability to determine what numerical operations are required to solve arithmetic problems without actually having to carry out the computations. It is similar to PY222, PY610, PY701, and PZ222.

EXAMPLE: Jane's father was 26 years old when she was born. Jane is now 8 years old. How old is her father now?

(A) subtract (C) add
(B) divide (D) multiply

SCALE STATISTICS:

MEAN = 10.80 ALPHA = 0.84 SAMPLE SIZE= 1465
ST.DEV= 3.56 ERR.MEAS= 1.41

ITEM STATISTICS:

ITEM INDEX	1	2	3	4	5	6	7	8	9	10
P	94	84	84	78	67	88	64	89	79	75
ADJ. P	94	84	84	79	67	89	65	92	83	81
BISERIAL	43	61	67	70	53	68	55	73	73	75
PERCENT NT	0	0	0	0	1	1	2	3	5	7
PAGE NO.	117	117	117	117	117	118	118	118	118	118
ITEM NO.	1	2	3	4	5	6	7	8	9	10

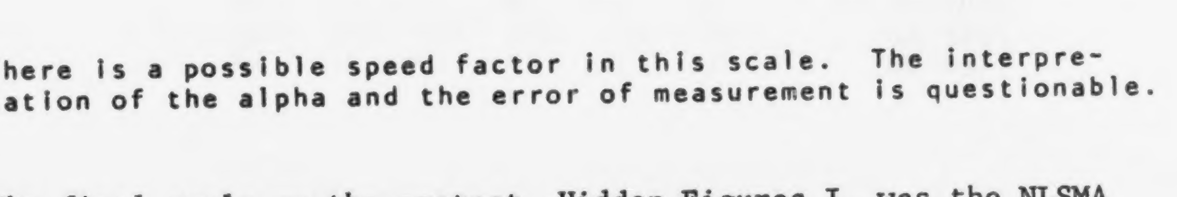
ITEM INDEX	11	12	13	14	15
P	60	54	65	61	37
ADJ. P	67	65	82	83	53
BISERIAL	53	41	82	61	35
PERCENT NT	11	16	21	26	31
PAGE NO.	119	119	119	119	119
ITEM NO.	11	12	13	14	15

There is a possible speed factor in this scale. The interpretation of the alpha and the error of measurement is questionable.

The final scale on the pretest, Hidden Figures I, was the NLSMA version of a test by the same name developed by ETS. A description of the test follows (NLSMA Reports, No. 4, 1968, p. 181).

PX218 HIDDEN FIGURES I (16 items; 15 minutes) This scale was patterned after the French Kit Form Cf-1, Part I. The original test has about twice as many distracting lines in each figure. It is designed to measure ability to keep a definite configuration in mind so as to make identification in spite of perceptual distractors. The task is to decide which of five geometrical figures is embedded in a complex pattern. It is similar to PX820, PY223, PY819, and PZ223.

EXAMPLE:



SCALE STATISTICS:

MEAN = 7.49
ST.DEV= 3.92
ALPHA = 0.81
ERR.MEAS= 1.70
SAMPLE SIZE= 1461

```
ITEM STATISTICS:

ITEM INDEX      1    2    3    4    5    6    7    8    9   10

P              72   52   58   61   51   36   39   55   53   60
ADJ. P         72   52   58   61   52   37   40   57   56   65
BISERIAL       41   50   55   51   52   38   56   58   62   60
PERCENT NT      0    0    0    0    1    2    3    4    5    7
PAGE NO.      105  105  105  105  106  106  106  106  107  107
ITEM NO.        1    2    3    4    5    6    7    8    9   10

ITEM INDEX     11   12   13   14   15   16

P              46   58   33   27   32   18
ADJ. P         51   66   40   36   43   27
BISERIAL       41   53   56   50   73   59
PERCENT NT     11   13   18   24   27   34
PAGE NO.      107  107  108  108  108  108
ITEM NO.       11   12   13   14   15   16


There is a possible speed factor in this scale.  The interpre-
tation of the alpha and the error of measurement is questionable.
```

A complete copy of the pretest is included in Appendix C.

The posttest consisted of two NLSMA tests, Working with Numbers and Five Dots and a set of five word problems developed specifically for the study.

The first scale in the posttest was an expanded version of the Working with Numbers Test used in the pretest. While the scale as described below consists of eight items, only the first three items are not duplicated in the pretest form of the scale. The posttest scale called Working with Numbers is X307 plus items 1, 2, and 3 of X313. This scale was labeled X307P to distinguish it from the pretest scale. The three items were included as problems numbered 11, 12, and 13 on the posttest scale of 15 items. The scale statistics for X313 are given below (NLSMA Reports, No. 4, 1968, p. 35).

X313 ANALYSIS I (8 items) This scale is intended to measure ability to analyze a problem situation and to apply knowledge appropriately. This scale has two items in common with X304 and five items in common with X307. It is the same as X611.

EXAMPLE: What number does ◇ stand for if
$3 \times 4 \times 5 = 12 \times$ ◇ is a true statement ?
(A) 20 (B) 0 (C) 3 (D) 4 (E) 5

SCALE STATISTICS:

MEAN = 3.49 ALPHA = 0.55 SAMPLE SIZE= 1332
ST.DEV= 1.85 ERR.MEAS= 1.24

ITEM STATISTICS:

ITEM INDEX	1	2	3	4	5	6	7	8
P	34	24	77	52	42	51	31	38
ADJ. P	34	25	77	52	42	51	31	38
BISERIAL	17	34	42	50	47	41	34	09
PERCENT NT	0	3	0	0	0	0	0	2
PAGE NO.	172	176	181	185	186	186	188	188
ITEM NO.	4	12	23	3	6	7	11	12

The second scale in the posttest was called Five Dots. The description of the scale and statistics are given here (NLSMA Reports, No. 4, 1968, p. 34).

X308 FIVE DOTS (19 items; 15 minutes) This scale is designed to measure ability to read a passage about a mathematical idea which is unfamiliar and to answer a series of questions about this idea. (No sample item has been included since the introductory explanation is so lengthy.) It is constructed to be logically parallel to X715.

SCALE STATISTICS:

MEAN = 10.50 ALPHA = 0.85 SAMPLE SIZE= 1330
ST.DEV= 4.68 ERR.MEAS= 1.78

ITEM STATISTICS:

ITEM INDEX	1	2	3	4	5	6	7	8	9	10
P	76	61	81	79	70	66	77	51	66	61
ADJ. P	76	61	81	79	70	66	77	51	67	61
BISERIAL	60	52	69	59	59	61	69	68	68	66
PERCENT NT	0	0	0	0	0	0	0	0	0	1
PAGE NO.	191	191	191	191	191	191	191	192	192	192
ITEM NO.	13	14	15	16	17	18	19	20	21	22

ITEM INDEX	11	12	13	14	15	16	17	18	19
P	62	41	33	34	35	37	34	46	41
ADJ. P	63	41	33	36	36	40	37	52	47
BISERIAL	68	52	54	43	48	61	49	47	46
PERCENT NT	1	1	2	4	5	6	9	11	13
PAGE NO.	192	192	192	193	193	193	193	193	193
ITEM NO.	23	24	25	26	27	28	29	30	31

The third scale on the posttest was a set of five arithmetic word problems. To obtain the problems for this scale, the author pilot-tested 150 word problems in three different forms in two classes at the fourth-, fifth-, and sixth-grade levels. Thirty of the problems were pilot tested in one class at the fourth-, sixth-, and eighth-grade levels to determine the relative difficulty of the problems. None of the classes or schools in which pilot testing was done was included in the main study. Five problems were selected for use on the posttest. The problems were chosen as being representative of their level of difficulty, in terms of the probability correct in the pilot test, and their complexity in terms of the number of steps required for solution. A copy of the problems in this scale, together with the complete posttest, is included in Appendix D.

The follow-up test consisted of three scales. The first scale, Letter Puzzles I was a NLSMA scale. Its description and statistics are given below (NLSMA Reports, No. 4, 1968, p. 53, 54).

X601 LETTER PUZZLES I (22 items; 8 minutes) This scale is designed to measure ability to handle a novel mathematical situation. It is composed of addition and subtraction problems in which some or all of the digits have been replaced by letters. The task is to determine all the missing digits. There are nine problems with 22 answers which are not independent. All items are completion items. This scale is similar to Y601.

EXAMPLE:

$$\begin{array}{r} ABB \\ +\ CB \\ \hline 174 \end{array}$$

A = _____ B = _____ C = _____

SCALE STATISTICS:

MEAN = 10.68 ALPHA = 0.90 SAMPLE SIZE= 1138
ST.DEV= 5.71 ERR.MEAS= 1.82

ITEM STATISTICS:

ITEM INDEX	1	2	3	4	5	6	7	8	9	10
P	81	75	72	79	43	33	34	74	75	49
ADJ. P	81	75	72	79	43	34	35	75	76	50
BISERIAL	63	60	52	53	66	74	70	68	69	68
PERCENT NT	0	0	0	0	1	1	1	1	1	2
PAGE NO.	268	268	268	268	268	268	268	268	268	268
ITEM NO.	1	2	3	4	5	6	7	8	9	10

(continued)

This scale is composed of nine two- and three-item clusters rather than 22 separate items. The alpha calculated with cluster scores rather than item scores is 0.83.

There is a possible speed factor in this scale. The interpretation of the alpha and the error of measurement is questionable.

ITEM INDEX	11	12	13	14	15	16	17	18	19	20
P	54	63	48	42	45	40	31	27	30	39
ADJ. P	55	66	50	44	51	45	39	34	38	57
BISERIAL	67	45	64	65	76	79	76	77	73	37
PERCENT NT	2	5	5	5	11	11	21	21	21	32
PAGE NO.	268	268	268	268	269	269	269	269	269	269
ITEM NO.	11	12	13	14	15	16	17	18	19	20

ITEM INDEX	21	22
P	20	15
ADJ. P	29	22
BISERIAL	58	53
PERCENT NT	32	32
PAGE NO.	269	269
ITEM NO.	21	22

The second scale on the follow-up test, Directions, was another of the NLSMA tests. Its description and statistics are as follows (NLSMA Reports, No. 4, 1968, p. 85).

X715 DIRECTIONS (24 items; 15 minutes) This scale is designed to measure ability to read a passage about a mathematical idea which is unfamiliar and to answer a series of questions about this idea. (No sample item has been included since the introductory explanation is so lengthy.) It is constructed to be logically parallel to X308.

SCALE STATISTICS:

MEAN = 15.16 ALPHA = 0.90 SAMPLE SIZE= 1111
ST.DEV= 6.14 ERR.MEAS= 1.93

ITEM STATISTICS:

ITEM INDEX	1	2	3	4	5	6	7	8	9	10
P	70	75	85	42	70	53	83	78	67	47
ADJ. P	70	75	85	42	70	53	83	78	67	47
BISERIAL	68	69	85	60	58	54	84	65	70	56
PERCENT NT	0	0	0	0	0	0	0	0	0	0
PAGE NO.	339	339	339	339	339	339	339	339	340	340
ITEM NO.	38	39	40	41	42	43	44	45	46	47

ITEM INDEX	11	12	13	14	15	16	17	18	19	20
P	83	69	59	38	46	49	71	72	64	65
ADJ. P	83	69	59	38	47	50	72	73	65	66
BISERIAL	85	70	56	38	70	53	72	84	70	77
PERCENT NT	0	0	0	0	0	0	1	1	2	2
PAGE NO.	340	340	340	340	340	340	341	341	341	341
ITEM NO.	48	49	50	51	52	53	54	55	56	57

ITEM INDEX	21	22	23	24
P	38	66	71	57
ADJ. P	39	69	77	64
BISERIAL	53	79	71	56
PERCENT NT	3	5	8	11
PAGE NO.	342	342	342	342
ITEM NO.	58	59	60	61

The third scale in the follow-up test was a parallel form of the five-item word-problem test given in the posttest. The problems were identical in every respect except that the names of the quantities in the problems were changed. The numerical values, the number of operations to solve each problem, and the order in which the problem statement was given were parallel. Each problem and its parallel had the same number of words. The problem-solving scale and the complete form of the follow-up test are included in Appendix E.

Analysis. The data were analyzed using covariance analysis. The assumptions upon which the analysis of covariance rests are (a) students were assigned to methods randomly, (b) achievement scores had a linear regression on the covariates within each method, (c) achievement scores had a normal distribution for students with the same ability in the same treatment, (d) the slope of the regression lines was the same for each method, (e) variances were homogeneous, and (f) achievement scores were a linear combination of independent components that included an overall mean, the treatment effect, a linear regression on the covariate, and an error term (Elashoff, 1968; Cochran, 1957).

In the present study, both groups were treated exactly the same except for treatment; since the booklets were handed out on a random basis to students in each class. The control group received no treatment. Concerning the assumptions stated above, there was no indication that the regressions used were not linear. The covariate scales were independent of treatment effects since they were administered before the treatments began. The achievement scores were assumed normal for each group of students. The sample sizes were such that this assumption seemed valid. The homogeneity of the regression lines were automatically checked by the computer program used to analyze the data. The variances were also assumed to be homogeneous, and the achievement scores were assumed to be a linear combination of the components listed above.

The analysis of covariance was run using the BMDX64 program (U.C.L.A., 1967) on an IBM 360/67 computer at the Stanford Computation Center by Mr. Ray Rees, Data Analysis Statistician for School Mathematics Study Group.

The covariates were the four pretest scales. The variates were the scale scores on the posttest and follow-up test.

CHAPTER III

RESULTS

This chapter begins with consideration of the reliability of the tests used, moves to the consideration of their inter-correlation and the tests of some assumptions underlying analysis of covariance, and then presents data on the treatment effects on the variate scales for each covariate.

In Chapter II the pretests, posttests, and follow-up tests were described and the Cronbach alpha for each, where available, was given. Table 2 presents for each test the alpha from the NLSMA data and the alpha computed from the data of the present study. For the Figure Classification Test, no alpha was available from the NLSMA data since the test was not a part of that study. Similiarly, the word-problem tests, P501 and F502, were developed specifically for this study and therefore were not comparable with any existing test. As can be seen from Table 2, the two sets of alphas are comparable although the sample size in the present study is considerably smaller than that in the NLSMA study.

The correlation coefficients among all criterion tests are presented in Table 3. All of the correlations are significant at the .01 level, which indicates that the tests apparently measured similar performances. The figures in parentheses give the number of students in each group indicated by the intersection of row and column.

The correlation coefficients of each of the criterion tests and the covariates (pretests) are presented in Table 4. The N in this table indicates the number of students for whom a complete set of data on all four

TABLE 2

Values for Cronbach's α and N for each Test from the NLSMA* Data and the Present Study

	Cronbach α			
	NLSMA Study		Present Study	
	α	N**	α	N
Pretests				
Figure Classification (X005)	--	--	.70	285
Working with Numbers (X307)	.68	1332	.59	285
Necessary Arithmetic Operations (PX217)	.84	1465	.84	285
Hidden Figures I (PX218)	.81	1461	.77	285
Posttests				
Working with Numbers (X307P)	.68/.55	1322	.65	265
Five Dots (X308)	.85	1330	.82	265
Word Problems (P501)	--	--	.43	265
Follow-up Tests				
Letter Puzzles I (X601)	.90	1138	.88	260
Directions (X715)	.90	1111	.79	265
Word Problems (F502)	--	--	.49	265

*National Longitudinal Study of Mathematical Abilities
**Total N for the present study

TABLE 3

Correlation Matrix for Posttest and Follow-up Test Scales

Scales		1	2	3	4	5	6
Working with Numbers (X307P)	1	1.000 (244)	0.548 (244)	0.426 (241)	0.397 (225)	0.435 (227)	0.500 (228)
Five Dots (X308)	2		1.000 (244)	0.385 (241)	0.475 (225)	0.507 (227)	0.434 (228)
Word Problems (P501)	3			1.000 (241)	0.240 (222)	0.336 (224)	0.523 (225)
Letter Puzzles (X601)	4				1.000 (233)	0.407 (232)	0.389 (233)
Directions (X715)	5					1.000 (235)	0.438 (235)
Word Problems (F502)	6						1.000 (236)

Note: The figures in parentheses indicate number of students in each computation.

$P(|r| > .19) < .001$ for $N = 244$
$P(|r| > .20) < .001$ for $N = 222$

TABLE 4

Correlation Coefficients for Each Posttest and Follow-up Test Scale with the Covariate Scales for all Students

Scales		1	2	3	4	N
Figure Classification (X005)	1					
Working with Numbers (X307)	2	-03				
Necessary Arithmetic Operations (PX217)	3	11	50			
Hidden Figures (PX218)	4	-05	49	35		
Working with Numbers (X307P)		-09	66	53	47	240
Five Dots (X308)		00	48	45	33	240
Word Problems (P501)		-16	32	48	26	237
Letter Puzzles (X601)		-01	45	42	32	229
Directions (X715)		02	42	40	36	231
Word Problems (F502)		-04	47	58	37	232

Note: Decimal points omitted.

$P(|r| > .13) < .05$; $P(|r| > .17) < .01$; $P(|r| > .22) < .001$ in a two-tailed test.

covariates and the criterion tests was obtained. For Word Problems (P501), the N of 237 indicates that 237 students had complete sets of data for all four covariate scales and the Word Problem (P501) scale. The figures indicating the significance level in Table 4 were compiled using the smallest value of N. Hence, for the larger N the values are at least at the same level of significance as those shown. Except for the Figure Classification Scale (X005), the correlation coefficients were all significant at the .001 level. The low or negative correlation of the Figure Classification Scale (X005) was surprising when compared with the other scales, since that scale was intended to be a measure of a student's ability to discover rules and explain things. It had been assumed that this scale would be positively and strongly correlated with the other scales used. The correlation coefficients in Table 4 also indicate that the Figure Classification Scale (X005) was not positively and strongly correlated with other pretest scales. Apparently the factors measured by each of the other pretest scales were quite different from those measured on the Figure Classification Scale (X005).

The correlation coefficients for each test for each treatment with each of the covariates are given in Table 5. The figure for N in each treatment group indicates the number of students in each group for whom there were complete sets of data for the four covariates and the criterion test shown. Several patterns are evident. First is the negative correlation of the Figure Classification Scale (X005) when compared with the other tests for students in both the Control and Wanted-Given Treatments. This strong negative relationship is not evident for students in The Productive Thinking Program. For boys and girls, the covariate with the greatest number of highly significant correlations is Necessary Arithmetic

TABLE 5

Correlations for Each Treatment Group by Variate Scale

		Boys					Girls				
		1	2	3	4	N	1	2	3	4	N
Productive Thinking Program											
Figure Classification	1										
Working with Numbers	2	36+					08				
Necessary Arith. Oper.	3	39	35				04	45			
Hidden Figures	4	32	55	29			-16	55	23		
Working with Numbers (X307P)		14	62**	30	34*	(40)	04	55**	55**	42**	(46)
Five Dots (X308)		36*	31	20	41*	(40)	32*	44**	63**	20	(46)
Word Problems (P501)		06	19	45**	31	(40)	04	42**	49**	14	(45)
Letter Puzzles (X601)		46**	43**	45**	42*	(38)	11	13	32*	11	(43)
Directions (X715)		23	51*	40*	34*	(38)	02	19	30*	35*	(43)
Word Problems (F502)		04	41*	58**	24	(39)	29	47**	62**	23	(43)
Modified Wanted-Given Program											
Figure Classification	1										
Working with Numbers	2	-22					-11				
Necessary Arith. Oper.	3	-02	48				06	71			
Hidden Figures	4	01	43	35			-24	58	54		
Working with Numbers (X307P)		-11	63**	51**	50**	(40)	-14	77**	70**	62**	(50)
Five Dots (X308)		-19	64*	49**	24	(40)	-07	62**	62**	48**	(50)
Word Problems (P501)		-24	24	54**	17	(40)	-38**	49**	40**	37**	(50)
Letter Puzzles (X601)		-06	62**	52**	39*	(39)	04	49**	51**	31*	(49)
Directions (X715)		06	54**	50**	39*	(40)	06	59**	60**	50**	(49)
Word Problems (F502)		-06	58**	62**	36*	(40)	04	60**	66**	52**	(49)
Control											
Figure Classification	1										
Working with Numbers	2	-43					-18				
Necessary Arith. Oper.	3	-29	34				05	46			
Hidden Figures	4	-39	52	43			-13	42	39		
Working with Numbers (X307P)		-53**	60**	49*	38	(27)	-33*	79**	53**	54**	(37)
Five Dots (X308)		-41*	31	18	40*	(27)	-08	30	49**	34*	(37)
Word Problems (P501)		-38	37	63**	40*	(25)	-32	25	44**	39*	(37)
Letter Puzzles (X601)		-40*	32	34	41*	(26)	-47**	54**	27	32	(34)
Directions (X715)		-33	07	27	12	(26)	-13	35*	26	35*	(35)
Word Problems (F502)		-44*	24	37	43*	(26)	-25	44*	52**	54**	(35)

Note: +Decimal points omitted.

*P < .05 on a two-tailed test.

**P < .01

Operations followed by Working with Numbers. The number of significant correlations ($P < .01$) of variate scales with covariate scales was highest for girls in the Modified Wanted-Given Program. The only covariate scale which was consistently highly correlated ($P \leq .01$) with Word Problems (P501) was Necessary Arithmetic Operations. With the exception of boys in the control group a significant correlation ($P < .01$) exists between the parallel form of the word problem scale (F502) and the same covariate, Necessary Arithmetic Operations. Differences in the number and pattern of correlation between criterion tests and covariates for the different treatment groups are also evident when examining Table 5. Overall, the greatest number of significant correlations ($P < .01$) was found in the data for students who received the Modified Wanted-Given Program.

The F-ratios in each row of the first four columns of Table 6 were computed using a one-way analysis of variance. The small differences in F-values in the first four columns were due to the variation in sample size in each cell as described earlier. Significant relationships were found to exist between the variate scales and two of the covariate scales, Figure Classification and Necessary Arithmetic Operations. The finding that significant differences did exist between covariate scales suggests that the use of covariance techniques for data analysis was appropriate.

The F-values for regression in Table 6 were computed using analysis of covariance. As may be seen in the Regression Colunn there were no significant differences due to regression. This indicates a lack of heterogeneity and thereby satisfies one of the basic assumptions of the covariance model. The complete analysis of variance tables for the covariance analysis are included in Appendix F.

TABLE 6

F-Ratios by Posttest and Follow-up Test Scale for Each Covariate and the Regression

Scale	Covariates				Regression
	Fig. Class.	Work. with Nos.	Nec. Arith. Oper.	Hidden Figures I	
Posttest					
Working with Numbers (X307P)	3.02*	1.49	3.19**	0.30	0.706
Five Dots (X308)	3.02*	1.49	3.19**	0.30	1.559
Word Problems (P501)	2.98*	1.46	3.06*	0.31	0.869
Follow-up Test					
Letter Puzzles I (X601)	4.17**	0.91	2.78*	0.20	1.083
Directions (X715)	4.26***	1.14	2.98*	0.22	0.802
Word Problems (F502)	4.30***	1.07	2.95*	0.22	1.423

*P < .05
**P < .01
***P < .001

The raw score means and standard deviations of criterion scales and covariates are shown by criterion test and treatment group criterion in Table 7. The adjusted score means for each treatment group are presented in Table 8. The effects of the covariance adjustment on the means may be seen by comparing corresponding entries from Tables 7 and 8.

The F-ratios for scale mean, treatment effect, sex, treatment x sex interaction, and covariates with each criterion test are shown in Table 9. One of the mean scores, the mean for Word Problems (F502), was not significantly different from zero. This was a surprising result. There were no significant differences due to treatment on any of the variate scales at the .05 level or beyond. Significance at the 0.1 level on the follow-up test, Word Problems (F502), was indicated for students in the Modified Wanted-Given Program. The single significant treatment x sex interaction ($P < .05$) was found for the posttest, Word Problems (P501). Boys in the Modified Wanted-Given Program gained significantly more than did the girls receiving that treatment.

The combination of covariates, labeled Total Covariates in Table 9, had a highly significant ($P < .005$) relation to the criterion scores. When considered separately, Covariate 3, Necessary Arithmetic Operations, was also found to have a significant ($P < .005$) relationship to each criterion score. None of the other covariates was significantly related to all the criterion scores. However, all covariates were significantly related (at least $P < .01$) to the posttest, Working with Numbers (X307P). The F-statistic for Word Problems (P501) and Covariate 1, Figure Classification, was significant ($P < .01$) for the posttest, but not for the follow-up test (F502). The scale Working with Numbers (X307) was not significantly related to Word Problems (P501) on the posttest, but was signifi-

TABLE 7

Raw Scores Means and Standard Deviations on Variate and Covariate Scale for Each Treatment Group

	Productive Thinking Program				Wanted-Given Program				Control				Total	
	Boys		Girls		Boys		Girls		Boys		Girls			
	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
	(N = 40)		(N = 46)		(N = 40)		(N = 50)		(N = 27)		(N = 37)		(N = 240)	
Working with Numbers (X307P)	6.07	2.63	6.41	2.67	6.57	3.03	6.34	3.17	6.04	2.34	6.86	2.71	6.40	2.79
Figure Classifications	39.75	11.10	44.30	9.68	38.47	13.35	42.84	9.30	40.48	8.55	45.92	8.80	42.09	10.53
Working with Numbers	4.65	2.13	4.35	2.17	4.45	2.25	4.72	2.75	5.11	1.78	5.54	2.08	4.76	2.27
Necessary Arithmetic Operations	7.15	3.84	8.09	3.60	7.92	3.91	8.92	3.91	7.81	3.85	10.27	3.63	8.39	3.88
Hidden Figures I	6.05	3.92	6.22	3.14	6.12	4.45	5.58	3.08	5.48	3.34	6.14	3.43	5.95	3.55
	(N = 40)		(N = 46)		(N = 40)		(N = 50)		(N = 27)		(N = 37)		(N = 240)	
Five Dots (X308)	10.95	3.79	10.50	4.22	11.47	4.62	10.96	4.92	11.56	3.94	11.27	3.92	11.07	4.27
Figure Classification	39.75	11.10	44.30	9.68	38.47	13.35	42.84	9.30	40.48	8.55	45.92	8.80	42.09	10.53
Working with Numbers	4.65	2.13	4.35	2.17	4.45	2.25	4.72	2.75	5.11	1.78	5.52	2.08	4.76	2.27
Necessary Arithmetic Operations	7.15	3.84	8.09	3.60	7.92	3.91	8.94	3.91	7.81	3.85	10.27	3.63	8.39	3.88
Hidden Figures I	6.05	3.92	6.22	3.14	6.12	4.45	5.58	3.08	5.48	3.34	6.14	3.43	5.95	3.55
	(N = 40)		(N = 45)		(N = 40)		(N = 50)		(N = 25)		(N = 37)		(N = 237)	
Word Problems (P501)	1.05	1.11	1.53	1.22	1.85	1.48	1.54	1.09	1.44	0.87	1.70	1.10	1.52	1.19
Figure Classification	39.75	11.10	44.24	9.78	38.47	13.35	42.84	9.30	40.24	8.86	45.92	8.80	42.05	10.59
Working with Numbers	4.65	2.13	4.44	2.09	4.45	2.25	4.72	2.75	5.24	1.74	5.54	2.08	4.79	2.26
Necessary Arithmetic Operations	7.15	3.84	8.13	3.62	7.92	3.91	8.94	3.91	8.16	3.77	10.27	3.63	8.44	3.87
Hidden Figures	6.05	3.92	6.31	3.11	6.12	4.45	5.58	3.08	5.52	3.37	6.14	3.43	5.97	3.55
	(N = 38)		(N = 43)		(N = 39)		(N = 49)		(N = 26)		(N = 34)		(N = 229)	
Letter Puzzles (X601)	10.24	4.11	9.95	4.77	10.33	5.48	9.76	5.97	10.19	4.63	11.94	5.36	10.34	5.14
Figure Classification	40.05	10.68	45.09	9.40	36.67	14.45	42.98	9.52	40.73	8.62	46.06	9.00	42.02	10.90
Working with Numbers	4.74	2.11	4.47	2.13	4.69	2.27	4.84	2.76	5.04	1.78	5.50	2.06	4.85	2.26
Necessary Arithmetic Operations	7.13	3.86	8.37	3.87	8.18	3.81	9.10	3.85	7.81	3.93	10.18	3.53	8.49	3.88
Hidden Figures	6.08	4.02	6.12	3.12	6.15	4.52	5.76	3.21	5.42	3.40	6.12	3.44	5.96	3.61
	(N = 38)		(N = 43)		(N = 40)		(N = 49)		(N = 26)		(N = 35)		(N = 231)	
Directions (X715)	13.21	4.66	12.77	4.07	13.80	4.66	13.65	4.11	13.69	3.70	13.63	4.28	13.44	4.25
Figure Classification	40.50	10.86	45.09	9.40	36.55	14.29	42.98	9.52	40.73	8.62	46.09	8.87	42.07	10.89
Working with Numbers	4.63	1.99	4.47	2.13	4.67	2.25	4.84	2.76	5.04	1.78	5.57	2.08	4.84	2.24
Necessary Arithmetic Operations	7.18	3.90	8.37	3.87	8.15	3.77	9.10	3.85	7.81	3.93	10.31	3.57	8.52	3.90
Hidden Figures	6.00	4.05	6.12	3.12	6.10	4.47	5.76	3.21	5.42	3.40	6.26	3.48	5.96	3.61
	(N = 39)		(N = 43)		(N = 40)		(N = 49)		(N = 26)		(N = 35)		(N = 232)	
Word Problems (F502)	1.33	1.22	1.56	1.22	1.95	1.58	1.86	1.14	1.42	1.36	1.89	1.43	1.69	1.33
Figure Classification	40.38	10.74	45.09	9.40	36.55	14.29	42.98	9.52	40.73	8.62	46.09	8.87	42.04	10.87
Working with Numbers	4.74	2.09	4.47	2.13	4.67	2.25	4.84	2.76	5.04	1.78	5.57	2.08	4.86	2.25
Necessary Arithmetic Operations	7.23	3.86	8.37	3.87	8.15	3.77	9.10	3.85	7.81	3.93	10.31	3.57	8.53	3.89
Hidden Figures	6.00	3.99	6.12	3.12	6.10	4.47	5.76	3.21	5.42	3.40	6.26	3.48	5.96	3.61

TABLE 8

Adjusted Means for Each Posttest and Follow-up Scale by Treatment Group

Scale	Productive Thinking Program		Modified Wanted-Given Program		Control		Total
	Boy	Girl	Boy	Girl	Boy	Girl	
Posttest							
Working with Numbers (X307P)	6.30	6.74	6.74	6.32	5.95	6.13	6.40
Five Dots (X308)	11.42	10.80	11.81	10.83	11.60	10.17	11.07
Word Problems (P501)	1.17	1.63	1.85	1.50	1.42	1.51	1.52
Follow-up Test							
Letter Puzzles (X601)	10.74	10.24	10.48	9.59	10.37	10.96	10.34
Directions (X715)	13.68	12.90	14.02	13.52	13.92	12.71	13.44
Word Problems (F502)	1.53	1.64	1.99	1.79	1.52	1.54	1.69

TABLE 9

F-Ratios for Mean, Treatment, Sex, Treatment-Sex Interaction and Covariates for Each Variate Scale

Scale	Mean	Treat.	Sex	Treat.-Sex	Total Cov.	Cov. 1	Cov. 2	Cov. 3	Cov. 4	N
Working with Numbers (X307P)	14.48***	1.35	0.07	1.08	63.83***	3.90*	69.38***	21.60***	6.96**	240
Five Dots (X308)	17.56***	0.26	4.11*	0.22	25.32***	0.04	18.78***	20.21***	1.47	240
Word Problems (P501)	9.99***	1.71	0.23	3.48*	22.49***	12.59***	1.46	41.96***	0.88	237
Letter Puzzles (X601)	7.07**	0.38	0.18	0.49	18.81***	0.05	14.71***	14.05***	1.82	229
Directions (X715)	36.02***	0.42	2.49	0.16	18.86***	0.33	9.93**	13.09***	5.61*	231
Word Problems (F502)	0.49	2.67+	0.03	0.51	36.14***	0.71	11.16***	51.20***	3.42+	232

+P < .10
*P < .05
**P < .01
***P < .001

cantly related ($P < .005$) to the parallel form (F502) in the follow-up test. This would seem to indicate an increase in the influence of computational skills on problem solving over time. The covariate scale Hidden Figures had a marginal statistical relationship ($P < .10$) with the follow-up Word Problem test (F502).

Mean gain scores for two sets of scales were tested for significance. First, an item analysis was run on Working with Numbers (X307P) after the three items which distinguished it from X307 were removed. The two tests were then identical. One test was given as one of the pretests. Four weeks later the other was given as a posttest. A t-test of the mean gain scores on the two tests was computed. No significant differences in mean gain scores were found for any of the three treatment groups on X307.

Mean gains for the Word Problem test (P501 and F502) were also tested for significance. Each of these tests consisted of five word problems. The only differences between items in each scale were in the names of things used in each problem. The computation required, the numbers given, the order of problems, and the number of words in each pair of problems were identical. There were no significant mean gains in terms of mean number correct in Word Problem scores over the seven-week period between the posttest and follow-up test. The data for the t-tests are presented in Table 10.

The lack of significant differences due to treatment is consistent with the null hypothesized, when one considers the findings from the standpoint of mean number correct on criterion tests. The Word-Problem tests were more difficult, on the average, than anticipated. The results from the analysis of mean number of problems correct were inconclusive, but favored the Modified Wanted-Given approach.

TABLE 10

Mean Gain Scores on Two Scales:
Working with Numbers and Word Problems

Scale	Treatment Group	N	Pretest	Posttest	Posttest-Pretest	t	df.
	1	87	4.48	4.78	0.30	1.28	86
Working with Numbers	2	93	4.59	4.81	0.22	1.04	92
X307P-X307	3	64	5.36	5.11	-0.25	-0.12	64
Total		244	4.75	4.88	0.13	0.97	243
Word Problems	1	77	1.36	1.51	0.15	1.00	76
F502-P501	2	89	1.73	1.83	0.10	0.75	88
	3	59	1.61	1.75	0.14	0.88	58
Total		225	1.57	1.70	0.12	1.51	224

To get some insight into whether students tried to follow procedures presented in the instructional treatments in solving each problem, regardless of whether their computations were right or wrong, all sets of word-problem tests (P501 and F502) were analyzed. A correct procedure was defined as a series of mathematical operations that would lead to a correct solution if the computation was correct at each step. Using the correct procedure was independent of whether or not mathematical sentences were used to "set up" the problem.

A new analysis of covariance was run to examine the effect of treatment on procedure for all students on the posttest scale Word Problems (P501). A significant difference was found ($P < .001$, $F_{2,199} = 12.04$) in favor of treatment groups 1 and 2. There was a significant treatment x sex interaction ($P < .05$, $F_{2,199} = 3.32$) in favor of boys on P501. The adjusted means indicated that students, in particular boys, who were given the Modified Wanted-Given Treatment were superior in achievement to those who were given _The Productive Thinking Program_ and the control group in that a greater number of the students were evidently using a procedure which would lead to a correct solution. A separate analysis of covariance was run, with the control group omitted, to test the significance of the differences in means of Modified Wanted-Given Treatment over Treatment 1. The results were significant ($P < .005$, $F_{1,146} = 6.23$) in favor of Treatment 2, the Modified Wanted-Given Treatment. Again, there was a significant ($P < .025$, $F_{1,146} = 6.03$) treatment x sex interaction in favor of boys.

The results of an analysis of covariance on the follow-up scale Word Problems (F502) indicated a significant difference in favor of the treatment groups ($P < .001$, $F_{2,207} = 7.87$) seven weeks after the main

TABLE 11

Adjusted Means on Word Problems (P501) for Procedures

Sex	Treatment Productive Thinking Program	Modified Wanted-Given Program	Control
Boys	2.07	2.93	1.60
Girls	2.28	2.31	1.75

TABLE 12

Adjusted Means on Word Problems (F502) for Procedures

Sex	Treatment Productive Thinking Program	Modified Wanted-Given Program	Control
Boys	2.68	2.96	2.07
Girls	2.40	2.66	1.95

study had ended. The adjusted means for the posttest are shown in Table 11 and the adjusted means for the follow-up test are shown in Table 12. In every case increase in adjusted mean scores from posttest to follow-up test was suggestive of some over-all effect, such as maturity or instruction. There was no significant treatment x sex interaction on F502 for all three treatment groups. There were no significant differences for the two treatment groups on the follow-up F502 when an analysis of covariance was run with the control group deleted. There was a marginal sex effect on F502 ($p < .10$, $F_{1,152} = 3.77$) in favor of boys. Evidently the differences between the two treatment groups in terms of using correct procedures decreased with the passage of time. It is interesting to compare the adjusted means for boys in both Treatment 1 and the control. On the posttest the adjusted means for girls were higher in these two groups. In the follow-up test the adjusted means for the boys were higher. The adjusted means for girls increased most for the Treatment 2 group, but the scores for boys in this group did not increase as much on posttest means as did the scores for boys in the other groups.

Taken as a whole, the results seem to indicate some degree of independence of computational skill and solution strategy in solving problems in arithmetic for fifth graders.

CHAPTER IV

SUMMARY AND DISCUSSION

Rationale. The study was intended to examine which of two approaches, both successful in their own area, was the better method of problem solving in mathematics. Researchers, such as Olton, Crutchfield, and others have argued for the existence of a general problem-solving ability and for the transferability of the general problem-solving skills developed through systematic instruction to specific disciplines such as arithmetic. The results of several studies have confirmed their position that students who receive instruction in The Productive Thinking Program apparently do become better problem solvers, while other studies have failed to find any significant differences. Failure to find significantly greater achievement levels in mathematics as transfer from training in The Productive Thinking Program, in studies to date, may have been due to the fact that the mathematical tasks used in some of these studies were too simple to demonstrate the full effects of the training. On the other hand, studies have shown that teaching students a wanted-given method of analyzing word problems in arithmetic can produce significant differences in achievement in solving one-step problems. This study attempted to provide a setting rich enough to test the hypotheses stated below.

Purpose. The purpose of this study was to test the hypotheses that: (a) there will be no significant differences in scores on posttest measures of problem-solving skills between students who receive training in general problem-solving techniques as presented in The Productive

Thinking Program and students who are taught more specific skills for solving problems in a total mathematics context using a Modified Wanted-Given Program approach; (b) students in either of the treatment groups will not achieve significantly higher scores on measures of problem-solving skills than students in control classes who receive no special instruction in problem solving; and (c) there will be no significant differences between the achievement of boys and girls on tests of problem-solving skills in the treatment groups.

Subjects. The subjects in this study were 261 fifth-grade students from four schools in the San Jose area of Northern California.

Treatments. The treatments consisted of: (a) The Productive Thinking Program Series One: General Problem Solving; (b) the author's Modified Wanted-Given Program; and (c) the control group. Treatments 1 and 2 were prepared in programmed booklet form. Each student completed one booklet each day for the 16 days of the instructional period. A pretest was administered before treatments began. A posttest was given the day following the end of the treatment period and a follow-up test was given seven weeks later. Students were assigned at random to treatments within classes in three of the schools. Students in the fourth school acted as a control. The results were analyzed using analysis of covariance.

Results. The results are first summarized in terms of mean number of problems correct on posttests and follow-up tests and later in terms of the procedure used by students to solve word problems in arithmetic. On the basis of the data from the present study, the results indicate that Hypothesis 1 cannot be rejected. No significant differences due to treatment, at the .05 level or above, were found on any of the posttests or follow-up tests. However, a difference at the .10 level favoring the Modified Wanted-Given Program was found on the follow-up test Word Problems (F502) indicating some degree of superiority of this treatment over The Productive Thinking Program, which was given seven weeks after the posttest was given.

Differences between means on parallel forms of the Word Problem test given as part of the posttest and seven weeks later as part of the follow-up test were not significant when subject to a t-test.

Perhaps the best indicator of a possible reason for the low-level significance of the Modified Wanted-Given Treatment on the Word Problem test is the difference in significance of covariates for Word Problems between the posttest and follow-up test. The covariate Working with Numbers did not have a significant relationship with the word problem scale on the posttest. The relationship was significant on the follow-up test. This finding tends to confirm an observation from the field; that is, the study was begun fairly early in the school year in an attempt to control any teacher-initiated instruction on problem solving. A side effect was that the students involved in the study apparently had had insufficient time to review their computational skills, so that their skills were actually below the level assumed by the writer when he wrote and tested the problems. While the results of the pilot tests gave no indication that the computational level was too difficult, close examination of the instructional booklets used in the Modified Wanted-Given Treatment indicated that many of the children had difficulty multiplying by two numbers, while others had difficulty with long division.

Hypothesis 2 cannot be rejected. There were no significant differences in achievement between treatment groups and the control group on any of the posttests or follow-up tests. While the adjusted means on the word problem test for the control group were somewhat lower than those for students in the Modified Wanted-Given Program, the scores of the control students were higher in one case than for those students in The Productive Thinking Program.

Of the six criterion tests, only one, Five Dots (X308), had a significant ($P < .05$) treatment x sex effect. In terms of the statement of Hypothesis 3, this was not considered sufficient evidence to warrant rejection of the hypothesis. This was the only one of the six possible tests for which significant sex effect was found, and since both Word Problem scales had low F values, Hypothesis 3 should not be rejected.

Two subsequent analyses, using covariance adjustment, of the procedures employed by students to solve problems in the Word Problem scales P501 revealed significant differences ($P < .001$) in favor of the treatment groups over the control, and a significant difference ($P < .005$) in favor of the Modified Wanted-Given Program over The Productive Thinking Program. Students who were given the Modified Wanted-Given Program treatment tended to use correct procedures to solve the mathematical word problems significantly more often than did their peers in either of the other two groups. This significant finding tends to confirm the validity of the instructional approach used in the Modified Wanted-Given Program, at least in the short term. The significance level in favor of the treatment groups over controls on the follow-up test F502 was also .001. However, the difference in favor of the Modified Wanted-

Given Program over The Productive Thinking Program on the follow-up test was not significant at the .10 level, when comparing the two treatment groups. There was however, a treatment x sex interaction ($P < .10$) in favor of boys in the Modified Wanted-Given Program. The effects of instruction and possibly maturation during the interval before the follow-up test, may account for these findings, since all adjusted means were higher on the follow-up test than on the posttest. This may indicate a greater degree of independence between computational facility and problem-solving ability in terms of knowing how to go about solving a problem than was previously suspected.

Limitations. The data analyzed and reported in the present study were collected from students in schools in a predominately white, middle-class socioeconomic area. All students in the study were fifth graders. Furthermore, the extremes in terms of gifted and handicapped, whether physical or educational, were not included in the data reported here. Any conclusions one might wish to draw must of necessity be made with the characteristics of the sample population clearly in mind. Perhaps the most important limitation was the finding that none of the pilot study groups reported the computational difficulty with the problems presented in the Modified Wanted-Given Program treatment that students in the main experiment reported. Had the study been delayed until after Christmas, thus permitting teachers to complete a thorough review of the basic computational skills in each class, significant effects in terms of the number of problems correct on each scale might have resulted.

Conclusions and Implications. On the basis of the data presented and within their limitations, it would appear that teaching problem solving in mathematics to fifth graders can best be done in a mathematical context using a wanted-given approach. One should not necessarily expect gains in the number of problems correctly solved, unless the teaching of problem-solving strategies is accompanied by instruction in the appropriate computational skills. The independence of these two factors seemed to be demonstrated when students in the experimental groups who were using a correct procedure to solve problems often failed to obtain significantly more mathematically correct answers, due to computational errors, than did the control group.

Providing systematic instruction in problem solving was more effective in this study in helping students to use correct procedures than not providing systematic instruction. This would seem to say that those who contend that students learn to solve problems by simply doing lots of problems may in fact be taking the long way home. Students may learn to solve problems that way, but students in the present study who received either treatment used appropriate strategies much more often ($P < .005$) than did the control group. This does not mean however, that following any "systematic" program will produce better results than simply having students solve lots of problems.

Suggestions for research. The question of the significance of computational difficulty in problem solving in mathematics is an open issue. Students in this study were sometimes observed to block when faced with a difficult computational exercise, and as a result, were unable to proceed at all toward the solution of the problem.

The effect of wording in problem statements is still largely unknown. It is clear, from the pilot study work done here, however, that simply reordering the sentences in a problem statement can significantly affect the difficulty level of a problem.

The effect of a student's age and the approach used in this study remains open. Torrance, as mentioned earlier, suggested that after the sixth grade students appear to reach a plateau in problem-solving achievement when the treatment vehicle is a series of programmed lessons such as those used herein. His studies were all concerned with general problem-solving ability. It would be most interesting if one could determine age boundaries in the effectiveness of a programmed approach to instruction in problem solving in mathematics if indeed such boundaries do exist.

A large number of other questions which relate to determining the relative difficulty of word problems remain to be answered. That is, what variables can one identify that influence the difficulty of word problems for students? Some of the variables that might be investigated are the number of words used in the problem statement, the syntactic difficulty of the sentences themselves, the sequence of problems presented, the number and type of operations required for solution, the memory load imposed by the problem solution sequence, the discrepancy between the order in which the numbers are given in the problem statement and the order in which they must be combined in binary operations for the solution of the problem, and many others. Some of these variables are currently being considered by researchers, and others will no doubt be considered in the near future.

The suggestion of the semi-independence of computational skill and procedural ability of students is very interesting. In most cases, in the author's experience, problem-solving tests are scored solely on the basis of the number of problems solved correctly, and no attempt is made to examine the procedures students may have been using. It may be that students are often taught how to proceed to solve problems correctly, but their new knowledge of how to solve a particular set of problems may be masked by the presence of computational errors. Computation and problem-solving ability may be highly correlated as many studies show, but the best measure of skill in problem solving will be one which brings to light the procedures used by students to solve problems as well as to give the correct answers. Studies of the sort reported here are needed to confirm or counter the differences noted above. If the findings and conclusions are supported by other studies, then work should begin on the development of new measures of problem solving--measures that evaluate both computation and procedures.

CHAPTER V

REVIEW OF RESEARCH IN PROBLEM SOLVING ON A COMPUTER-BASED TELETYPE

The first attempt at IMSSS to apply the linear regression model described in Suppes, Jerman, and Brian (1968) to the analysis of word problems in arithmetic was reported in Suppes, Loftus, and Jerman (1969). A second study on word-problem variables using regression models was reported by Loftus (1970). What is desired from models of this sort is that they identify and define structural variables that are suspected to account for differences in the level of difficulty for a variety of types of word problems in terms of the structure of the problem itself. That is, one would like to attach weights to the several objectively defined variables suspected of contributing to the relative difficulty of the problems, and then to use the estimates of the variables in predicting the relative difficulty of a large number of problems.

A discussion of the regression model is not given here, since it has been described in detail in the three references cited above. Rather, the focus of this chapter is on a review of the variables used in the first two of the previous studies, on their definitions, and on how well each accounted for the variance observed in data from a new set of problems not solved at a teletype but in a regular classroom using paper and pencil. Some additional variables which were developed and used later are also discussed.

The variables used in Suppes, Loftus, and Jerman (1969, p. 7) were as follows.

X_1 Operations: the minimum number of different operations required to reach the correct solution;

X_2 Steps: the minimum number of steps required to reach the correct solution;

X_3 Length: the number of words in the problem;

X_4 Sequential: assigned a value of 1 if the problem is of the same type (i.e., can be solved by the same operation(s)) as the problem that preceded it, and 0 otherwise;

X_5 Verbal-cue: assigned a value of 1 if the problem contains a verbal cue to the operation(s) required to solve the problem, and 0 otherwise;

X_6 Conversion: assigned a value of 1 if a conversion of units is required to solve the problem, and 0 otherwise.

The 1969 study reported the results of the goodness of fit for 68 word problems which 27 gifted fifth-grade students solved as part of their daily CAI program. The regression equation for all 68 items was

$$z_i = -7.36 + .87X^*_{i1} + .18X_{i2} + .02X^*_{i3} + .26X_{i5} + 1.42X^*_{i6} \, .$$

The multiple R was .67 with standard error of 1.75, and an R^2 of .45. The χ^2 value for the 68 items was 555.76, indicating a rather poor fit for the model.

---

* indicates significance at the .05 level.

Perhaps one of the variables in the above set which requires a description for clarification is X_5, verbal cue. The following words were considered cues.

Word	Cue for
and	addition
left	subtraction
each	multiplication
average or each	division

The conversion variable was defined for the situation in which the student was to recall a fact from memory, such as 1 week = 7 days. This conversion, or substitution of units was a 0, 1 variable.

Several people in the Institute became interested in formulating variables and testing their goodness of fit on the data from the Suppes, Loftus, and Jerman (1969) study. In particular, Dr. Barbara Searle formulated and tested several variables using the data set from the 1969 study. The set of variables she tested consisted of a mixture of some of the above variables and some new variables which she formulated. The variables were defined as follows:

Operations: the minimum number of operations required to reach a correct solution (values range 1-4).

Steps: the minimum number of binary operations, steps, needed to reach a solution (value range 1-7).

Length: the number of words in the problem (value range 7-51).

Conversion: this factor is present if a conversion is required and the equivalent units are not given in the problem (a 0, 1 variable).

Verbal cue: the cue for each operation is as follows.

Operation	Cue words
Addition:	added, altogether, gained
Subtraction:	how much less, lost, left
Multiplication:	each
Division:	average

If a cue word was present the value was 1, otherwise 0.

Order: If the steps to solution were in order as given in the problem statement, the value was 1, otherwise 0.

Formula: If knowledge of a formula was required, the value was 1, otherwise 0.

Average: If the problem statement contained the word "average" the value was 1, otherwise 0.

Addition: If the problem requires addition, the value is 1, otherwise 0.

Subtraction: If the problem requires subtraction, the value is 1, otherwise 0.

Multiplication: If the problem requires multiplication, the value is 1, otherwise 0.

Division: If the problem requires division, the value is 1, otherwise 0.

Sequence: If the problem is in unusual order, the value is 1, otherwise 0.

Three of the problems in the original set of 68 were deleted due to their very high X^2 values. The following variables were tested on the data from the remaining 65 problems.

Of the 16 variables in the expanded set, 12 were entered by the stepwise regression program, BMD02R. The value of the multiple R was .820. Three additional variables were formulated after studying the weights of the variables, their contribution to the total R^2 and their definitions. Two of these, S_1 and S_2, were sequential variables; the third was a memory variable. The definitions for the three additional variables follow.

Memory (M) is defined as the sum of:

C the number of conversions + knowledge of formulas,
D the number of numerals in the problem statement,
O the number of different operations.

S_1 is defined as the number of displacements of order of operations in successive problems.

Examples:

$$\left.\begin{matrix}3 + 4\\3 - 4\end{matrix}\right\} S_1 = 1 \qquad \left.\begin{matrix}3 + 5\\4 + 6\end{matrix}\right\} S_1 = 0$$

$$\left.\begin{matrix}(3 + 4) \times 2\\(3 + 4) \div 2\end{matrix}\right\} S_1 = 1 \qquad \left.\begin{matrix}(3 + 4) \times 2\\(3 \times 4) + 2\end{matrix}\right\} S_1 = 2$$

S_2 is defined as the number of displacements between order of operations required and that given in the problem statement itself.

The results of the stepwise regression using all the variables which contributed at least .01 to the variance in K^2 on the data from the set of 65 original problems are presented in Table 13.

TABLE 13

Results of Stepwise Regression on 65 Word Problems

A. Three additional variables alone

Step	Variable	R	Increase in R^2	Last reg. coefficient
1	19 memor	.491	.241	0.106
2	17 s1	.501	.010	0.073
3	18 s2	.513	.012	0.087

B. Over-all variables

Step	Variable	R	Increase in R^2	Last reg. coefficient
1	4 opers	.657	.431	.199
2	8 vblcu	.697	.055	.262
3	15 divis*	.729	.046	.299
4	6 lenth	.761	.047	.018
5	10 formu*	.785	.038	.879
6	17 s1	.805	.032	.130
7	7 convr	.825	.032	.534
8	18 s2	.835	.016	.127
9	14 multi*	.838	.005	-.090
10	19 memor	.840	.004	-.048
11	12 add*	.841	.002	.063
12	16 seque	.842	.001	.053

*Variables suggested by Dr. Searle.

In Part A in Table 13 the last three variables are presented alone. The total R for the three was .51. Part B shows the results of the stepwise regression using all variables. Only those variables are considered important which contributed to at least .01 increase in R^2. Although the value of the multiple-R term for all 12 variables was fairly high, .842, the number of variables entered did not increase with the addition of the three new ones. Rather, the three new variables entered instead of others and the resulting R increased from .820 to 842 ($R^2 = 0.709$). A total of 19 variables had now been formulated and tested on the original data. In addition to those described above, the following were formulated and tested.

Operations 2: The sum of the following.

1. The number of different operations.
2. Add 4 if one of the operations is division.
   Add 2 if one of the operations is multiplication.
   Add 1 if one of the operations is addition.

Order 2: The sum of the following.

1. S_1
2. Verbal cue necessary to establish a new order. One point for each direct cue missing for each step.

Recall: The sum of the following.

1. One count for a formula to be recalled and one count for each step in the formula, e.g., $A = 2\ell + 2w = 3$.
2. One count for each conversion to be recalled and used.
3. One count for each fact recalled and used from a previous problem.

Verbal cue 2: The set of cues was expanded. However, one count was given for each cue present in the problem.

Addition: added, altogether, gained, total.
Subtraction: how much less, lost, left
how much larger ... than
how much smaller ... than
how much greater ... than
how much further... than
Multiplication: each, times
Division: average

Distractors: This variable was defined as one count for each verbal cue which was not a cue for an operation, but a distractor. For example, if the word "average" was used, but multiplication rather than division was the required operation.

A complete list of variables, by number, follows.

Variable	Name	Variable	Name
1.	P(correct-observed)	12.	Multiplication
2.	Operations	13.	Division
3.	Steps	14.	Sequence
4.	Length	15.	S_1
5.	Conversions	16.	S_2
6.	Verbal Cue	17.	Memory
7.	Order	18.	Operations-2
8.	Formula	19.	Order-2
9.	Average	20.	Recall
10.	Addition	21.	Verbal Cue-2
11.	Subtraction	22.	Distractor Cue

It appears from observing the values of R and the increase in R^2 with each step that there is relatively little gain after the tenth step in either Table 13 or Table 14. In fact if one were to adopt the rule to consider as important only those variables whose contribution to the increase in R^2 was .01 or greater, then the first eight variables in Table 13 and the first nine variables in Table 14 would comprise the set of variables of interest in each case.

It is most interesting to examine the order of entry of the variables in each case, and at this stage, it is impossible to explain the differences. About all we can say to the reader is "behold." The following list ranks the variables under discussion for ease of comparison.

From Table 13			From Table 14		
Step		R	Step		R
1	Operations	.657	1	Operations	.657
2	Verbal Cue	.697	2	Conversions	.702
3	Division	.729	3	Length	.740
4	Length	.761	4	Order 2	.780
5	Former	.785	5	Division	.805
6	S_1	.805	6	S_2	.821
7	Conversions	.825	7	Order	.829
8	S_2	.835	8	Memory	.835
			9	Distractor Cues	.841

At this point, note that the variables which appear to be the most robust are length, division, S_2 the internal sequence variable, and conversions. Memory and distractor cues may or may not play important roles in subsequent analyses for students working at teletype terminals.

TABLE 14

Summary Table for the Stepwise Regression on 65 Items

Step Num.	Name	Variable Ent.	Variable Rem.	Multiple R	Multiple R^2	Increase in R^2	F Value For Del	Last Reg. Coefficients
1	Oper	2		0.6566	0.4311	0.4311	47.732	0.1824
2	Convr	5		0.7021	0.4929	0.0618	7.559	0.5517
3	Lenth	4		0.7405	0.5483	0.0554	7.488	0.0274
4	Order 2	19		0.7804	0.6090	0.0607	9.321	0.1427
5	Divis	13*		0.8050	0.6480	0.0390	6.516	0.3841
6	S_2	16		0.8213	0.6745	0.0265	4.749	0.1791
7	Order	7		0.8288	0.6869	0.0124	2.232	-0.1401
8	Memory	17		0.8349	0.6971	0.0102	1.872	-0.0839
9	Dist	22		0.8409	0.7071	0.0101	1.984	-0.3052
10	Recal	20		0.8444	0.7130	0.0059	1.105	0.1455
11	Verblc	6		0.8457	0.7152	0.0022	0.414	-0.1035
12	Verbl 2	21		0.8477	0.7186	0.0034	0.629	0.1350
13	S_1	15		0.8498	0.7222	0.0036	0.641	0.0482
14	Aver	9*		0.8504	0.7232	0.0010	0.202	-0.1551
15	Oper 2	18		0.8513	0.7247	0.0015	0.261	-0.0328
16	Sub	11*		0.8520	0.7259	0.0012	0.197	-0.0607
17	Seq	14		0.8527	0.7271	0.0012	0.210	0.0865
18	Steps	3		0.8529	0.7274	0.0003	0.063	0.0306

*Variables suggested by Dr. Searle.

CHAPTER VI

APPLICATION OF STRUCTURAL VARIABLES TO WORD PROBLEMS SOLVED IN A PAPER-AND-PENCIL CONTEXT

The research reported in Chapters I-IV was conducted entirely off line, in a paper-and-pencil mode. In the course of the treatment, students worked through several sets of word problems that were evaluated as tests of their achievement, but were not labeled as tests. Other sets of word problems were labeled as tests and were included as part of the posttest and follow-up test instruments. The following is a report of an analysis, using the variables defined in Chapter V, of one set of word problems that tested the generalizability of the model to off-line situations.

Twenty-nine problems were selected for analysis. Of these problems, 19 were solved by students (N = 20) using paper and pencil. The 19 problems were part of the instructional treatment provided by the Modified Wanted-Given Program, and the remaining 10 problems came from the two test scales P501 and F502 (N = 161). Eleven of the variables used in the earlier analyses were tested on the word problems solved off line to see if their order of entry in the stepwise regression was at all similar. The variables selected for testing were the following.

2. Operations 2
3. Order 2
4. Recall
8. Length
9. Verbal Cue (as defined by Dr. Searle)
10. Conversion

5. S_1
6. Memory
7. S_2
11. Formula
12. Division (as defined by Dr. Searle)

The results of the stepwise regression using these variables are summarized in Table 15. As can be seen, Length, the number of words in the statement of the problem, entered first followed by Memory, S_2, S_1, and Verbal Cue. The total R after nine steps was .77, $R^2 = .595$. This was somewhat surprising and gratifying. The variables accounting for most of the variance on line were also effective, though at a lower level, for accounting for most of the variance off line where the students did all the required computation by hand. The observed and predicted probability correct for each item is shown in Table 16.

The conclusions must be recognized as artificial, however. While the students did in fact solve all 29 problems, they solved only the first 19 at one sitting. The last 10 problems were solved 5 at a time, as post-test and follow-up tests, seven weeks apart. The sequential variable is not correct, therefore, for problems 19 and 25. Otherwise the variable values are the same as if students had taken the entire set of problems at one time. However, the results cannot be interpreted as being constant for students who would solve all 29 items at one sitting. Only the values of the variables are the same.

TABLE 15

Summary Table for 29 Word Problems Administered in

Paper-and-pencil Format, 11 Original Variables

Step		Variable		Multiple		Increase	F Value	Last Reg.
Num.	Name	Ent.	Rem.	R	R^2	in R^2	For Del	Coefficients
1	Lenth	8		0.5487	0.3011	0.3011	11.633	0.0207
2	Memor	6		0.5938	0.3526	0.0515	2.069	0.1588
3	S_2	7		0.6672	0.4455	0.0926	4.167	-0.3048
4	S_1	5		0.6977	0.4868	0.0416	1.949	-0.1379
5	Vblcu	9		0.7399	0.5475	0.0607	3.082	0.5784
6	Oper 2	2		0.7549	0.5699	0.0224	1.148	0.0962
7	Recal	4		0.7661	0.5869	0.0170	0.863	-0.2790
8	Convr	10		0.7708	0.5941	0.0072	0.356	-0.0797
9	Divis	12		0.7713	0.5949	0.0008	0.038	-0.0857

TABLE 16

Observed and Predicted Probability Correct for 29 Word Problems, 11 Original Variables

Case Num.	Observed	Predicted	Residual
1	0.2660	0.2116	0.0544
2	0.2410	0.4491	-0.2081
3	0.3100	0.3437	-0.0337
4	0.6200	0.3835	0.2365
5	0.2300	0.4580	-0.2280
6	0.3720	0.4066	-0.0346
7	0.0100	0.1297	0.3703
8	0.7770	0.6099	0.1671
9	0.1740	0.2201	-0.0461
10	0.5600	0.5600	0.0000
11	0.5380	0.3542	0.1838
12	0.1150	0.1681	-0.0531
13	0.0870	0.0870	0.0000
14	0.1000	0.1692	-0.0692
15	0.0710	0.1180	-0.0470
16	0.2940	0.2013	0.0927
17	0.1660	0.2939	-0.1279
18	0.0100	0.5333	-0.0333
19	0.0590	0.0989	-0.0399
20	0.4380	0.3783	0.0597
21	0.2440	0.3652	-0.1212
22	0.4540	0.3325	0.1215
23	0.3550	0.3967	-0.0417
24	0.1050	0.1440	-0.0390
25	0.3410	0.3783	-0.0373
26	0.3770	0.3652	0.0118
27	0.5170	0.3744	0.1426
28	0.3870	0.3967	-0.0097
29	0.1610	0.1228	0.0382

In an attempt to improve the fit, two new variables were defined. The first, Verbal Cue-1 (No. 13), was a redefinition of Dr. Searle's variable. It was essentially the definition used in Verbal Cue-2 except that it was a 0-1 variable rather than a frequency variable as is Verbal Cue 2. The second new variable (No. 14) was a combination of Verbal Cue 1 and indirect cues, such as "in all," for addition, "short of ...," for subtraction and "per ...," for multiplication. The results of the regression analysis on the 29 problems using 14 variables are shown in the summary Table 17. The increase in the value of R due to the addition of the two additional variables was small, almost .03 (.771 to .799). The observed and predicted scores for each problem are shown in Table 18.

TABLE 17

Summary Table for 29 Word Problems Administered in

Paper-and-pencil Form, 13 Original Variables

Step Num.	Variable Ent. Rem.	Multiple R	R^2	Increase in R^2	F Value For Del.	Last Reg. Coefficients
1	8	0.5487	0.3011	0.3011	11.633	0.0270
2	6	0.5938	0.3526	0.0515	2.069	0.2115
3	7	0.6672	0.4452	0.0926	4.167	-0.2184
4	5	0.6977	0.4868	0.0416	1.949	-0.1845
5	9	0.7399	0.5475	0.0607	3.082	0.5859
6	14	0.7688	0.5911	0.0436	2.350	-0.0761
7	2	0.7872	0.6197	0.0286	1.577	0.0329
8	10	0.7964	0.6343	0.0146	0.797	-0.1165
9	13	0.7974	0.6359	0.0016	0.080	-0.1453
10	12	0.7986	0.6378	0.0019	0.100	0.1586
11	4	0.7991	0.6386	0.0008	0.034	-0.1257
12	3	0.7993	0.6389	0.0003	0.013	-0.0085

TABLE 18

Observed and Predicted Probability Correct for 29 Word Problems, 13 Original Variables

Case Num.	Observed	Predicted	Residual
1	0.2660	0.2333	0.0327
2	0.2410	0.4579	-0.2169
3	0.3100	0.2820	0.0280
4	0.6200	0.4387	0.1813
5	0.2300	0.3734	-0.1434
6	0.3720	0.3539	0.0181
7	0.5000	0.1335	0.3665
8	0.7770	0.5531	0.2239
9	0.1740	0.1841	-0.0101
10	0.5600	0.5600	0.0000
11	0.5380	0.3984	0.1396
12	0.1150	0.1788	-0.0638
13	0.0870	0.0870	0.0000
14	0.1000	0.1862	-0.0862
15	0.0710	0.1292	-0.0582
16	0.2940	0.2633	0.0307
17	0.1660	0.3195	-0.1535
18	0.5000	0.5403	-0.0403
19	0.0590	0.0773	-0.0183
20	0.4380	0.4312	0.0068
21	0.2440	0.2974	-0.0534
22	0.4540	0.3782	0.0758
23	0.3550	0.4277	-0.0727
24	0.1050	0.1267	-0.0217
25	0.3410	0.4312	-0.0902
26	0.3770	0.2974	0.0796
27	0.5170	0.4429	0.0741
28	0.3870	0.4277	-0.0407
29	0.1610	0.0999	0.0611

Two variables described earlier, Verbal Cue-2 (No. 15) and a distractor variable (No. 16), were added to the regression. The results of the regression using all 15 variables are shown in Tables 19 and 20. The value of R increased from .799 with 13 variables to .834 with 15 variables.

The practicality of taking into account more than four or five variables when writing word problems is impossible, even if they were able to account for a larger portion of the variance than that indicated in Table 19. Clearly, the fit of the variables selected and tried thus far was less than satisfactory. The fact that the off-line students performed all computations by hand led to the definition of four new computational variables. These followed the work reported in Suppes and Morningstar (in press, Chapter III). The variables were:

17. EXMC. A 0-1 variable. A count of 1 was assigned for each multiplication exercise required in the solution of the problem. Zero was assigned if multiplication was not required.
18. NOMC2. A count of 1 was assigned each time a regrouping occurred in each multiplication exercise in the problem. For example:

$$\begin{array}{r} 38 \\ \times\ 5 \\ \hline 190 \end{array} \qquad \text{NOMC} = 2$$

$$\begin{array}{r} 38 \\ \times\ 25 \\ \hline 190 \\ 76 \\ \hline 950 \end{array} \qquad \text{NOMC} = 3$$

TABLE 19

Summary Table for 29 Word Problems

in Paper-and-pencil Format, 15 Original Variables

Step Num.		Variable Ent. Rem.	Multiple R	Multiple R^2	Increase in R^2	F Value For Del.	Last Reg. Coefficients
1		8	0.5487	0.3011	0.3011	11.633	0.0408
2		6	0.5938	0.3526	0.0515	2.069	0.1712
3		7	0.6672	0.4452	0.0926	4.167	-0.3204
4		15	0.7180	0.5155	0.0704	3.488	-0.2837
5		5	0.7465	0.5573	0.0417	2.167	-0.2402
6		12	0.7667	0.5878	0.0306	1.629	-0.2402
7		10	0.7798	0.6081	0.0203	1.089	-0.1611
8		9	0.7912	0.6260	0.0179	0.954	0.7250
9	Vbl 3	3	0.8077	0.6524	0.0264	1.445	0.0946
10	Vbl 4	4	0.8169	0.6673	0.0150	0.807	0.4544
11		14	0.8300	0.6889	0.0216	1.180	-0.1815
12		13	0.8332	0.6942	0.0053	0.285	0.1214
13		12	0.8332	0.6942	0.0000	0.004	
14		16	0.8341	0.6957	0.0015	0.079	-0.0520
15	Vbl 2	2	0.8343	0.6961	0.0003	0.014	0.0119

TABLE 20

Observed and Predicted (Correct) for 15 Variables

Case Num.	Observed	Predicted	Residual
1	0.2660	0.2421	0.0239
2	0.2410	0.5279	-0.2869
3	0.3100	0.2968	0.0132
4	0.6200	0.4379	0.1821
5	0.2300	0.3357	-0.1057
6	0.3720	0.2751	0.0969
7	0.5000	0.1966	0.3034
8	0.7770	0.6401	0.1369
9	0.1740	0.2482	-0.0742
10	0.5600	0.5435	0.0165
11	0.5380	0.4196	0.1184
12	0.1150	0.1912	-0.0762
13	0.0870	0.0870	0.0000
14	0.1000	0.1062	-0.0062
15	0.0710	0.1067	-0.0357
16	0.2940	0.2847	0.0093
17	0.1660	0.4012	-0.2352
18	0.5000	0.5462	-0.0462
19	0.0590	0.0676	-0.0086
20	0.4380	0.3839	0.0541
21	0.2440	0.2646	-0.0206
22	0.4540	0.3912	0.0628
23	0.3550	0.3854	-0.0304
24	0.1050	0.1450	-0.0400
25	0.3410	0.3839	-0.0429
26	0.3770	0.2646	0.1124
27	0.5170	0.4822	0.0348
28	0.3870	0.3854	0.0016
29	0.1610	0.1048	0.0562

19. COLC. For this variable a count of 1 was given for each column and a count of 1 was given for each regrouping in the largest addition and subtraction exercise in the problem. If no addition or subtraction was required, a count of 0 was given.

20. QUOT. A count of 1 was given for each digit in the quotient if division was required and 0 otherwise.

The set of 19 variables was coded for use on the set of 29 problems which had been rearranged from highest to lowest in order of probability correct. One additional problem (P(correct) = .50) was added to bring the total number of problems up to 30. The set of 30 problems was being prepared for administration to several groups of new students for replication of the earlier analyses.

A regression was run on the 30 problems using all 19 variables. The summary of this analysis is presented in Table 21. The value of R after the first five steps was 0.9315, $R^2 = 0.86769$. This is a surprisingly good fit for just five variables. Perhaps even more suprising was the strength of the computational variables indicated by their point of entry into the regression program. Of the first five variables which entered the regression, three were computational variables--NOMC a multiplication variable, QUOT a division variable, and COLC an addition and subtraction variable. The variable LENTH which accounted for the number of words in the problem statement entered first and the distractor variable DIST entered on the fourth step of the regression. The cognitive variables such as memory and order did not enter as soon or in the same order for students solving problems at a CAI terminal. The observed and predicted probability correct for each problem are presented in Table 22.

TABLE 21

Summary of Regression Analysis of 30 Problems

Using 19 Variables

Step Num.	Variable Ent.	Rem.	Multiple R	R^2	Increase in R^2	F Value For Del.	Last Reg. Coefficients
1		8	0.6684	0.4468	0.4468	22.609	0.0179
2		18	0.8395	0.7048	0.2580	23.586	0.2419
3		20	0.9129	0.8334	0.1286	20.067	0.2323
4		16	0.9254	0.8564	0.0230	4.013	0.1333
5		19	0.9315	0.8677	0.0113	2.055	0.0684
6		5	0.9350	0.8742	0.0065	1.189	0.0421
7		7	0.9400	0.8836	0.0094	1.761	-0.1679
8	VBL4	4	0.9428	0.8889	0.0053	1.012	-0.3726
9		17	0.9462	0.8953	0.0064	1.206	0.0896
10		6	0.9470	0.8968	0.0015	0.303	-0.0902
11	VBL2	2	0.9477	0.8981	0.0013	0.220	0.0600
12	VBL3	3	0.9487	0.9000	0.0019	0.329	0.0580
13		15	0.9499	0.9023	0.0023	0.362	-0.1434
14		14	0.9532	0.9086	0.0063	1.052	0.0880
15		10	0.9541	0.9103	0.0017	0.260	0.0565
16		9	0.9542	0.9105	0.0002	0.038	-0.0711
17		13	0.9543	0.9107	0.0002	0.010	0.0298

TABLE 22

Observed and Predicted Probability Correct

for Each Problem

Case Num.	Observed	Predicted	Residual
1	0.7770	0.6645	0.1125
2	0.6200	0.5727	0.0473
3	0.5600	0.5600	0.0000
4	0.5380	0.3721	0.1659
5	0.5170	0.4585	0.0585
6	0.5000	0.5510	-0.0510
7	0.4540	0.4457	0.0083
8	0.4380	0.5774	-0.1394
9	0.3870	0.3254	0.0616
10	0.3770	0.1840	0.1930
11	0.3720	0.6047	-0.2327
12	0.3550	0.3345	0.0205
13	0.3410	0.4647	-0.1237
14	0.3100	0.1917	0.1183
15	0.2940	0.2953	-0.0013
16	0.2660	0.2168	0.0492
17	0.2440	0.2957	-0.0517
18	0.2410	0.2779	-0.0369
19	0.2300	0.2740	-0.0440
20	0.1740	0.1845	-0.0105
21	0.1660	0.1317	0.0343
22	0.1610	0.1497	0.0112
23	0.1150	0.1427	-0.0277
24	0.1050	0.1533	-0.0483
25	0.1000	0.1082	-0.0082
26	0.0870	0.0870	0.0000
27	0.0710	0.0813	-0.0103
28	0.0590	0.0584	0.0006
29	0.0100	0.0153	-0.0053
30	0.0100	0.0098	0.0002

Only the first five variables contributed to an increase in R^2 of one per cent or more. The regression equation after the fifth step was

$$z_i = -.73 + .02X_2 + .19X_3 + .22X_4 + .03X_5 + .23X_6 \ .$$

Table 23 presents the regression coefficients, standard errors of regression coefficients, and computed T-values for each of the five variables. Only one of the variables, COLC, failed to reach significance at the .05 level.

TABLE 23

Regression Coefficients, Standard Errors of Regression Coefficients, and Computed T-values

Variable	Regression Coefficient	Standard Error	Computed T-value
X_2 length	.01955	.00626	3.10703**
X_3 distractor	.18769	.08924	2.1003*
X_4 multiplication	.22076	.02705	8.1537***
X_5 addition-subtraction	.02819	.01966	1.4847
X_6 division	.22983	.05979	3.8429***

$^{*}P < .05$ $^{**}P < .01$ $^{***}P < .001$

The correlation matrix for the variables is shown in Table 24. Variable one is the observed percentage correct and variable seven is the transgenerated dependent variable. The mean and standard deviations for each variable are given in Table 25. The summary for the five-step regression analysis is given in Table 26.

TABLE 24

Correlation Matrix for Each Variable

Variable Number	1	2	3	4	5	6	7
1	1.000	-0.637	-0.533	-0.413	0.001	-0.454	-0.926
2		1.000	0.610	-0.007	-0.205	0.606	0.668
3			1.000	-0.145	-0.147	0.605	0.543
4				1.000	-0.090	-0.226	0.503
5					1.000	-0.118	-0.092
						1.000	0.566
							1.000

TABLE 25

Means and Standard Deviations for Each Variable

Variable		Mean	Standard Deviation
P(C)	1	29.59667	19.55739
Lenth	2	35.73333	9.63447
Distr	3	0.33333	0.66089
Nomc	4	1.46667	1.67607
Colc	5	2.13333	2.25501
Quot	6	0.43333	1.00630
	7	0.51014	0.58148

TABLE 26

Summary of Five-step Regression Analysis

Multiple R	0.9315
Std. Error of Est.	0.2325

Analysis of Variance:

	DF	Sum of Squares	Mean Square	F-Ratio
Regression	5	8.508	1.702	31.485
Residual	24	1.297	0.054	

Variables in Equation: (Constant = -0.73442)

Variable	Coefficient	Std. Error	F to Remove
Lenth 2	0.01955	0.00626	9.7476 (2)
Distr 3	0.18769	0.08924	4.4233 (2)
Nomc 4	0.22076	0.02705	66.6153 (2)
Colc 5	0.02819	0.01966	2.0545 (2)
Quot 6	0.22983	0.05979	14.7775 (2)

In summary, the linear regression model described in this paper has given a surprisingly good account of the difficulty level of a somewhat artificially arranged set of verbal problems for fifth-grade students. Five variables were found to account for almost 87 per cent of the variance in the observed probability correct. The variable which accounted for most of the variance was NOMC (32 per cent) a multiplication variable, followed by QUOT (23 per cent) a division variable, then LENTH (21 per cent) the number of words in the problem statement, DISTR (11 per cent) the verbal distractor variable, and finally COLC (1 per cent) the addition-subtraction variable.

Before one tries to apply these findings to a real-life situation, however, a good deal of replication needs to be done to either confirm or deny the findings reported here. For example, which variables enter first for students at different grade levels or for students with different backgrounds or skill levels. Further, variables for students solving problems on a CAI system that eliminates the need for computational facility should be compared with the variables that were found to account for most of the variance in this study, i.e., for students required to do their own computations. Perhaps we can then get a clear picture of the distinction between levels of cognitive skills. It may be that stepwise regression techniques will provide a more suitable vehicle for determining whether aptitude-interaction factors do exist.

One indication of the differential effect among the several variables for on-line and off-line problem solving is indicated by the data shown in Table 27. The set of variables found so effective in predicting the probability correct for the set of 30 word problems was tested on the data from the Suppes, Loftus, and Jerman (1969) study.

As can be seen from Table 27, the overall fit is not as good and the order of entry of the variables is quite different from that shown in Table 21.

The first five variables from each analysis are shown below. The multiple R after five steps is

Step	From Table 21	From Table 27
1	Lenth	Oper 2
2	Nomc	Lenth
3	Quo	Ordr 2
4	Distr	Recal
5	Colc 2	S_2
R	.932	.747

also given. The data from Table 21 were based on paper-and-pencil work and the data from Table 27 on CAI work. It is clear from a study of the tables that the same variables do not enter in the same order or account for the same amount of increase in R^2 in each case. The variable Lenth does enter early in each case, however, and seems to be a consistent early entrant.

It is unfortunate that the intact set of problems analyzed in Table 21 was never run on the CAI system so that the variables entered both on line and off line could be analyzed. Many of the thirty problems in the set were included in the problem-solving strand, but upon entry the problem set was split and items were scattered throughout the strand according to another set of variables all without the author's knowledge. This made comparison for purposes of the present study practically impossible. Nevertheless, it does appear that the variables affecting item difficulty

TABLE 27

Computational Variables Run on Suppes, Loftus and Jerman (1969) Data

Step Num.	Name	Variable Ent. Rem.	Multiple		Increase in R^2	F Value For Del	Last Reg. Coefficients
			R	R^2			
1	Oper 2	8	0.5722	0.3274	0.3274	30.674	0.0427
2	Lenth	2	0.6539	0.4276	0.1002	10.844	0.0266
3	Ordr 2	9	0.7070	0.4999	0.0723	8.809	0.1543
4	Recall	10	0.7367	0.5427	0.0429	5.641	0.2526
5	S_2	6	0.7466	0.5574	0.0147	1.949	0.1276
6	Nomc	12	0.7607	0.5787	0.0213	2.923	-0.5043
7	Memory	7	0.7725	0.5968	0.0181	2.555	-0.0440
8	Colc	15	0.7768	0.6034	0.0067	0.938	0.1212
9	Divis	3	0.7782	0.6056	0.0029	0.310	0.1165
10	Vblc 2	11	0.7796	0.6078	0.0029	0.304	-0.0576
11	Quo	16	0.7803	0.6089	0.0011	0.133	0.0111
12	S_1	5	0.7806	0.6093	0.0005	0.080	-0.0268
13	Seq	4	0.7810	0.6100	0.0006	0.071	0.0458

are different when comparing problems that do not require computation with problems that do require computation.

It is too early to attempt an exploration of the observed differences in the entry points of the variables used. More work needs to be done. However, preliminary evidence suggests that one must not jump to conclusions about which variables are the most important contributers to the difficulty of verbal problems in arithmetic without first considering the tools at the disposal of the student who is to solve the problem.

References

Balow, I. Reading and computation ability as determinants of problem solving. The Arithmetic Teacher, 1964, 11, 18-22.

Burns, P., & Yonally, J. Does the order of presentation of numerical data in multi-steps arithmetic problems affect their difficulty? School Science and Mathematics, 1964, 64, 267-70.

Campbell, D., & Stanley, T. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Covington, M., & Crutchfield, R. Facilitation of creative problem solving. Programmed Instruction, 1965, 4, 3-5, 10.

Covington, M., Crutchfield, R., & Davies, L. The productive thinking program, series one: General problem solving. Berkeley, California: Educational Innovation, 1966.

Cronbach, L. The meanings of problems, in Arithmetic 1948. Supplementary Educational Monographs, Chicago: University of Chicago Press, 1948, 66.

Cochran, W. Analysis of covariance: Its nature and uses. Biometrics, 1957, 13, 261-281.

Crutchfield, R. Sensitization and activation of cognitive skills. In J. Bruner (Ed.), Learning about learning: A conference report. Washington, D. C.: U.S. Office of Education, 1966. Pp. 64-70.

Davis, G. Current status of research and theory in human problem solving. Psychological Bulletin, 1966, 66, 36-54.

Dodson, J. Characteristics of successful insightful problem solvers. Unpublished doctoral dissertation. University of Georgia, 1970.

Duncan, E. R., Capps, L. R., Dalcioni, M. P., Quast, W. G., & Zweng, M. Modern school mathematics structure and use. Teacher's annotated edition. Boston: Houghton Mifflin, 1967.

Dye, N., & Very, P. Growth changes in factorial structure by age and sex. Genetic Psychology Monographs, 1968, 78, 55-88.

Educational Testing Service. I-3 Figure Classifications. Princeton, New Jersey: 1963.

Eicholz, R., O'Daffer, P., Brumfiel, C., Shanks, M., & Fleenor, C. School mathematics I. Reading, Mass.: Addison-Wesley, 1967.

Eisenstadt, J. Problem-solving ability of creative and noncreative college students. Journal of Consulting Psychology, 1966, 30, 81-83.

Elashoff, J. Analysis of covariance. Research memorandum No. 34. August, 1968. Stanford University Center for Research and Development in Teaching.

Evans, E. Measuring the ability of students to respond in creative mathematical situations at the late elementary and early junior high school level. (Doctoral dissertation, University of Michigan.) Dissertation Abstracts 25: 1965, 12, 7107-7108 (Abstract).

Fisher, R., & Yates, F. Statistical tables for biological, agricultural and medical research. London: Oliver and Boyd, 1957.

Gangler, J. An experimental study of the effects of participation and motivation on the problem solving ability of college freshmen. (Doctoral dissertation, Columbia University.) Dissertation Abstracts 28: 1967, 2157B. 1967, 5 (Abstract).

Goals for School Mathematics. The Report of the Cambridge Conference on School Mathematics. Boston: Houghton Mifflin, 1963.

Gorman, C. A critical analysis of research on written problems in elementary school mathematics. (Doctoral dissertation, University of Pittsburgh.) Dissertation Abstracts 28: 1968, 12, 4818A-4819A (Abstract).

Henderson, P., & Pingry, R. Problem solving in mathematics. In The Learning of Mathematics, Its Theory and Practice. Twenty-first yearbook of the National Council of Teachers of Mathematics. Washington, D. C.: NCTM, 1953.

Hill, S. A study of logical abilities of children. Unpublished doctoral dissertation, Stanford University, 1960.

James, G., & James, R. Mathematics dictionary (Multilingual Edition). Princeton, Van Nostrand, 1968.

Jonsson, H. Interaction of test anxiety and test difficulty in mathematics problem-solving performance. (Doctoral dissertation, University of California, Berkeley.) Dissertation Abstracts 26: 1966, 7, 3757-3758 (Abstract).

Kellmer, P., & McKenzie, I. Teaching method and rigidity in problem solving. British Journal of Educational Psychology, 1965, 35, 50-59.

Kilpatrick, J. Problem-solving and creative behavior in mathematics. In J. Wilson and R. Carry (Eds.), Studies in Mathematics. Vol. 19. Stanford, Calif. School Mathematics Study Group, 1969.

Klein, P., & Kellner, H. Creativity in a two-choice probability situation. Journal of General Psychology, 1967, 76, 193-200.

Kromer, K. The teaching of elementary school mathematics. Boston: Allyn and Bacon, 1966.

Loftus, E. J. F. An analysis of the structural variables that determine problem-solving difficulty on a computer-based teletype. Technical Report No. 162, December 18, 1970, Stanford University, Institute for Mathematical Studies in the Social Sciences.

Martin, M. Reading comprehension, abstract verbal reasoning, and computation as factors in arithmetic problem solving. (Doctoral dissertation, State University of Iowa.) Dissertation Abstracts 24: 1964, 11, 4547-48 (Abstract).

Mendelsohn, G., & Griswold, B. Differential use of incidental stimuli in problem solving as a function of creativity. Journal of Abnormal and Social Psychology, 1964, 68, 431-436.

Mitchell, C. Problem analysis and problem-solving in arithmetic. Elementary School Journal, 1932, 32, 464-466.

Neimark, E., & Lewis, N. The development of logical problem-solving strategies. Child Development, 1967, 38, 107-117.

O'Brien, T., & Shapiro, B. The development of logical thinking in children. American Educational Research Journal, 1968, 5, 531-542.

Olton, R., Wardrop, J., Covington, M., Goodwin, W., Crutchfield, R., Klausmeier, H., & Ronda, T. The development of productive thinking skills in fifth-grade children. Technical Report No. 34, 1967, University of Wisconsin, Wisconsin Research and Development Center for Cognitive Learning.

Olton, R., & Crutchfield, R. S. Developing skills of productive thinking. In P. Musson, J. Langer, & M. V. Covington (Eds.), New directions in developmental psychology. New York: Holt, Reinhart and Winston (in press).

Olton, R. A self-instructional program for developing productive thinking skills in fifth- and sixth-grade children. Journal of Creative Behavior, 1969, 3, 16-25.

Polya, G. How to solve it. (2nd ed.) Garden City, N. Y.: Doubleday, 1957.

Prouse, H. Creativity in school mathematics. Mathematics Teacher, 1967, 60, 876-879.

Ray, W. Complex tasks for use in human problem-solving research. Psychological Bulletin, 1955, 52, 134-149.

Riedesel, C. Problem solving: Some suggestions from research. The Arithmetic Teacher, 1969, 16, 54-55.

Ripple, R., & Dacey, J. The facility of problem solving and verbal creativity by exposure to programmed instruction. Psychology in the Schools, 1967, 4, 240-245.

Sheehan, J. Patterns of sex differences in learning mathematical problem-solving. The Journal of Experimental Education, 1968, 36, 84-87.

Stern, C. Acquisition of problem-solving strategies by young children, and its relation to verbalization. Journal of Educational Psychology, 1967, 58, 245-252.

Stern, C., & Keisler, E. Acquisition of problem-solving strategies by young children, and its relation to mental age. American Educational Research Journal, 1967, 4, 1-12.

Stull, L. Auditory assistance of reading as a factor in intermediate-grade pupil's interpretations of verbal arithmetic problems. (Doctoral dissertation, The Pennsylvania State University.) Dissertation Abstracts 25: 1965, 12, 7113 (Abstract).

Suppes, P., Jerman, M., & Brian, D. Computer-assisted instruction: The 1965-66 Stanford arithmetic program. New York: Academic Press, 1968.

Suppes, P., Loftus, E., & Jerman, M. Problem-solving on a computer-based teletype. Educational Studies in Mathematics, 1969, 2, 1-15.

Suppes, P., & Morningstar, M. Computer-assisted instruction at Stanford, 1966-1968: Data, models, and evaluation. New York: Academic Press, in press.

Suydam, M., & Redesel, C. Interpretative study of research and development in elementary school mathematics. Three volumes. Final Report, 1969, The Pennsylvania State University, Project No. 8-0586, Grant No. OEG-0-9-480586-1352(010).

Tate, M., & Stanier, B. Errors in judgment of good and poor problem solvers. Journal of Experimental Education, 1964, 32, 371-376.

Thiels, C. A comparison of three instructional methods in problem solving. Research on the Foundations of American Education. Washington, D. C.: American Educational Research Association, 1939.

Thompson, E. Readability and accessory remarks: Factors in problem solving in arithmetic. (Doctoral dissertation, Stanford University.) Dissertation Abstracts 28: 1968, 7, 2464A-2465A (Abstract).

Treffinger, D., & Ripple, R. Developing creative problem solving abilities and related attitudes through programmed instruction. Journal of Creative Behavior, 1969, 3, 105-110.

Torrance, G. Understanding the fourth-grade slump in creative thinking. Final Report on Cooperative Research Project 994, 1967, University of Georgia, Georgia Study of Creative Behavior (ERIC No. ED 018 273).

University of California at Los Angeles, Health Sciences Computing Facility. BMD Biomedical Computer Programs. Berkeley: University of California Press, 1967.

Van Engen, H. Twentieth century mathematics for the elementary school. The Arithmetic Teacher, 1959, 6, 71-76.

Vanderlinde, L. Does the study of quantitative vocabulary improve problem-solving? Elementary School Journal, 1964, 65, 143-152.

Very, P. Differential factor structures in mathematical ability. Genetic Psychology Monographs, 1967, 75, 169-207.

Weidelin, I. A synthesis of two factor analyses of problem solving in mathematics. Malmo, Sweden: Malmo School of Education, 1966.

West, C., & Loree, M. Selectivity, redundancy, and contiquity as factors which influence the difficulty of an achievement test. Journal of Experimental Education, 1968, 36, 89-93.

Williams, M., & McCreight, R. Shall we move the question? The Arithmetic Teacher, 1965, 12, 418-421.

Wilson, J. The role of structure in verbal problem solving in arithmetic: An analytical and experimental comparison of three problem-solving programs. (Doctoral dissertation, Syracuse University.) Ann Arbor, Mich.: University Microfilms, 1964, No. 65-3445.

Wilson, J. Generality of heuristics as an instructional variable. (Doctoral dissertation, Stanford University.) Dissertation Abstracts 28: 1967, 1, 2575A (Abstract).

Wilson, J., Cahen, L., & Begle, E. (Eds.). Description and statistical properties of X-population scales. Stanford, Calif.: School Mathematics Study Group, 1968.

Wittrock, M. Replacement and nonreplacement strategies in children's problem solving. Journal of Educational Psychology, 1967, 58, 69-74.

Zweng, M. A reaction to the role of structure in verbal problem solving. The Arithmetic Teacher, 1968, 15, 251-253.

List of Appendices

Appendix A	Sample pages from the Modified Wanted-Given Program Instructional Treatment
Appendix B	Instruction set for teaching aides
Appendix C	Pretest
Appendix D	Posttest
Appendix E	Follow-up test
Appendix F	Analysis of variance tables for the covariate analysis for each variate scale

The appendices listed above are not included with this report. Copies may be obtained by writing to the author at School of Education, Pennsylvania State University.

(Continued from inside front cover)

96 R. C. Atkinson, J. W. Breisford, and R. M. Shiffrin. Multi-process models for memory with applications to a continuous presentation task. April 13, 1966. (*J. math. Psychol.*, 1967, *4*, 277-300).

97 P. Suppes and E. Crothers. Some remarks on stimulus-response theories of language learning. June 12, 1966.

98 R. Bjork. All-or-none subprocesses in the learning of complex sequences. (*J. math. Psychol.*, 1968, *1*, 182-195).

99 E. Gammon. The statistical determination of linguistic units. July 1, 1966.

100 P. Suppes, L. Hyman, and M. Jerman. Linear structural models for response and latency performance in arithmetic. (In J. P. Hill (ed.), *Minnesota Symposia on Child Psychology*. Minneapolis, Minn.: 1967. Pp. 160-200).

101 J. L. Young. Effects of intervals between reinforcements and test trials in paired-associate learning. August 1, 1966.

102 H. A. Wilson. An investigation of linguistic unit size in memory processes. August 3, 1966.

103 J. T. Townsend. Choice behavior in a cued-recognition task. August 8, 1966.

104 W. H. Batchelder. A mathematical analysis of multi-level verbal learning. August 9, 1966.

105 H. A. Taylor. The observing response in a cued psychophysical task. August 10, 1966.

106 R. A. Bjork. Learning and short-term retention of paired associates in relation to specific sequences of interpresentation intervals. August 11, 1966.

107 R. C. Atkinson and R. M. Shiffrin. Some Two-process models for memory. September 30, 1966.

108 P. Suppes and C. Ihrke. Accelerated program in elementary-school mathematics--the third year. January 30, 1967.

109 P. Suppes and I. Rosenthal-Hill. Concept formation by kindergarten children in a card-sorting task. February 27, 1967.

110 R. C. Atkinson and R. M. Shiffrin. Human memory: a proposed system and its control processes. March 21, 1967.

111 Theodore S. Rodgers. Linguistic considerations in the design of the Stanford computer-based curriculum in initial reading. June 1, 1967.

112 Jack M. Knutson. Spelling drills using a computer-assisted instructional system. June 30, 1967.

113 R. C. Atkinson. Instruction in initial reading under computer control: the Stanford Project. July 14, 1967.

114 J. W. Breisford, Jr. and R. C. Atkinson. Recall of paired-associates as a function of overt and covert rehearsal procedures. July 21, 1967.

115 J. H. Stelzer. Some results concerning subjective probability structures with semiorders. August 1, 1967

116 D. E. Rumelhart. The effects of interpresentation intervals on performance in a continuous paired-associate task. August 11, 1967.

117 E. J. Fishman, L. Keller, and R. E. Atkinson. Massed vs. distributed practice in computerized spelling drills. August 18, 1967.

118 G. J. Groen. An investigation of some counting algorithms for simple addition problems. August 21, 1967.

119 H. A. Wilson and R. C. Atkinson. Computer-based instruction in initial reading: a progress report on the Stanford Project. August 25, 1967.

120 F. S. Roberts and P. Suppes. Some problems in the geometry of visual perception. August 31, 1967. (*Synthese*, 1967, *17*, 173-201)

121 D. Jamison. Bayesian decisions under total and partial ignorance. D. Jamison and J. Kozielecki. Subjective probabilities under total uncertainty. September 4, 1967.

122 R. C. Atkinson. Computerized instruction and the learning process. September 15, 1967.

123 W. K. Estes. Outline of a theory of punishment. October 1, 1967.

124 T. S. Rodgers. Measuring vocabulary difficulty: An analysis of item variables in learning Russian-English and Japanese-English vocabulary parts. December 18, 1967.

125 W. K. Estes. Reinforcement in human learning. December 20, 1967.

126 G. L. Wolford, D. L. Wessel, W. K. Estes. Further evidence concerning scanning and sampling assumptions of visual detection models. January 31, 1968.

127 R. C. Atkinson and R. M. Shiffrin. Some speculations on storage and retrieval processes in long-term memory. February 2, 1968.

128 John Holmgren. Visual detection with imperfect recognition. March 29, 1968.

129 Lucille B. Mlodnosky. The Frostig and the Bender Gestalt as predictors of reading achievement. April 12, 1968.

130 P. Suppes. Some theoretical models for mathematics learning. April 15, 1968. (*Journal of Research and Development in Education*, 1967, *1*, 5-22)

131 G. M. Olson. Learning and retention in a continuous recognition task. May 15, 1968.

132 Ruth Norene Hartley. An investigation of list types and cues to facilitate initial reading vocabulary acquisition. May 29, 1968.

133 P. Suppes. Stimulus-response theory of finite automata. June 19, 1968.

134 N. Moler and P. Suppes. Quantifier-free axioms for constructive plane geometry. June 20, 1968. (In J. C. H. Gerretsen and F. Oort (Eds.), *Compositio Mathematica*. Vol. 20. Groningen, The Netherlands: Wolters-Noordhoff, 1968. Pp. 143-152.)

135 W. K. Estes and D. P. Horst. Latency as a function of number or response alternatives in paired-associate learning. July 1, 1968.

136 M. Schlag-Rey and P. Suppes. High-order dimensions in concept identification. July 2, 1968. (*Psychom. Sci.*, 1968, *11*, 141-142)

137 R. M. Shiffrin. Search and retrieval processes in long-term memory. August 15, 1968.

138 R. D. Freund, G. R. Loftus, and R.C. Atkinson. Applications of multiprocess models for memory to continuous recognition tasks. December 18, 1968.

139 R. C. Atkinson. Information delay in human learning. December 18, 1968.

140 R. C. Atkinson, J. E. Holmgren, and J. F. Juola. Processing time as influenced by the number of elements in the visual display. March 14, 1969.

141 P. Suppes, E. F. Loftus, and M. Jerman. Problem-solving on a computer-based teletype. March 25, 1969.

142 P. Suppes and Mona Morningstar. Evaluation of three computer-assisted instruction programs. May 2, 1969.

143 P. Suppes. On the problems of using mathematics in the development of the social sciences. May 12, 1969.

144 Z. Domotor. Probabilistic relational structures and their applications. May 14, 1969.

145 R. C. Atkinson and T. D. Wickens. Human memory and the concept of reinforcement. May 20, 1969.

146 R. J. Titiev. Some model-theoretic results in measurement theory. May 22, 1969.

147 P. Suppes. Measurement: Problems of theory and application. June 12, 1969.

148 P. Suppes and C. Ihrke. Accelerated program in elementary-school mathematics--the fourth year. August 7, 1969.

149 D. Rundus and R.C. Atkinson. Rehearsal in free recall: A procedure for direct observation. August 12, 1969.

150 P. Suppes and S. Feldman. Young children's comprehension of logical connectives. October 15, 1969.

(Continued on back cover)

(Continued from inside back cover)

151 Joaquim H. Laubsch. An adaptive teaching system for optimal item allocation. November 14, 1969.
152 Roberta L. Klatzky and Richard C. Atkinson. Memory scans based on alternative test stimulus representations. November 25, 1969.
153 John E. Holmgren. Response latency as an indicant of information processing in visual search tasks. March 16, 1970.
154 Patrick Suppes. Probabilistic grammars for natural languages. May 15, 1970.
155 E. Gammon. A syntactical analysis of some first-grade readers. June 22, 1970.
156 Kenneth N. Wexler. An automaton analysis of the learning of a nimiature system of Japanese. July 24, 1970.
157 R. C. Atkinson and J.A. Paulson. An approach to the psychology of instruction. August 14, 1970.
158 R.C. Atkinson, J.D. Fletcher, H.C. Chetin, and C.M. Stauffer. Instruction in initial reading under computer control: the Stanford project. August 13, 1970.
159 Dewey J. Rundus. An analysis of rehearsal processes in free recall. August 21, 1970.
160 R.L. Klatzky, J.F. Juola, and R.C. Atkinson. Test stimulus representation and experimental context effects in memory scanning.
161 William A. Rottmayer. A formal theory of perception. November 13, 1970.
162 Elizabeth Jane Fishman Loftus. An analysis of the structural variables that determine problem-solving difficulty on a computer-based teletype. December 18, 1970.
163 Joseph A. Van Campen. Towards the automatic generation of programmed foreign-language instructional materials. January 11, 1971.
164 Jamesine Friend and R.C. Atkinson. Computer-assisted instruction in programming: AID. January 25, 1971.
165 Lawrence James Hubert. A formal model for the perceptual processing of geometric configurations. February 19. 1971.
166 J. F. Juola, I.S. Fischler, C.T. Wood, and R.C. Atkinson. Recognition time for information stored in long-term memory.
167 R.L. Klatzky and R.C. Atkinson. Specialization of the cerebral hemispheres in scanning for information in short-term memory.
168 J.D. Fletcher and R.C. Atkinson. An evaluation of the Stanford CAI program in initial reading (grades K through 3). March 12, 1971.
169 James F. Juola and R.C. Atkinson. Memory scanning for words versus categories.
170 Ira S. Fischler and James F. Juola. Effects of repeated tests on recognition time for information in long-term memory.
171 Patrick Suppes. Semantics of context-free fragments of natural languages. March 30, 1971.
172 Jamesine Friend. Instruct coders' manual. May 1, 1971.
173 R.C. Atkinson and R.M. Shiffrin. The control processes of short-term memory. April 19, 1971.
174 Patrick Suppes. Computer-assisted instruction at Stanford. May 19, 1971.
175 D. Jamison, J.D. Fletcher, P. Suppes and R.C.Atkinson. Cost and performance of computer-assisted instruction for compensatory education.
176 Joseph Offir. Some mathematical models of individual differences in learning and performance. June 28, 1971.
177 Richard C. Atkinson and James F. Juola. Factors influencing speed and accuracy of word recognition. August 12, 1971.
178 P. Suppes, A. Goldberg, G. Kanz, B. Searle and C. Stauffer. Teacher's handbook for CAI courses. September 1, 1971.
179 Adele Goldberg. A generalized instructional system for elementary mathematical logic. October 11, 1971.
180 Max Jerman. Instruction in problem solving and an analysis of structural variables that contribute to problem-solving difficulty. November 12, 1971.